



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

보건학 석사 학위논문

마커 유전자 조합을 이용한
비결핵 마이코박테리아 동정 및
분류 시스템 연구

**NON-TUBERCULOUS MYCOBACTERIA
CLASSIFICATION SYSTEM BASED ON
MULTIPLE MARKER GENES**

2015년 8월

서울대학교 보건대학원
보건학과 생명정보학 전공
조 광 훈

마커 유전자 조합을 이용한 비결핵 마이코박테리아 동정 및 분류 시스템 연구

지도교수 손 현 석

이 논문을 보건학 석사학위논문으로 제출함

2015년 5월

서울대학교 보건대학원

보건학과 생명정보학 전공

조 광 훈

조광훈의 석사학위논문을 인준함

2015년 6월

위 원 장	조 성 일	(인)
부 위 원 장	정 해 원	(인)
위 원	손 현 석	(인)

국 문 초 록

마커 유전자 조합을 이용한 비결핵 마이코박테리아 동정 및

분류 시스템 연구

서울대학교 보건대학원 보건학과

생명정보학 전공

조 광 훈

인간이 살아가는 환경에는 수많은 균들이 존재하고 인간에게 유익한 균도 있고 해가 되는 균도 있다. 그 중에서 비결핵 마이코박테리아균은 마이코박테리아 속(genus)에서 결핵균과 나병균을 제외한 균을 말한다. 비결핵 마이코박테리아는 결핵균과 구분 없이 인식되다가 1980년대 AIDS의 확산과 함께 크게 확산된 박테리아균이다. 자연계에 감염보유 숙주가 없는 결핵균과는 달리 비결핵 마이코박테리아는 물과 토양이 있는 주변 환경에서 발견된다. 비결핵 마이코박테리아는 주로 호흡기, 경부 림프선, 피부 등에 감염을 일으키며, 이 균에 의한 질환 중 90% 이상이 호흡기 질환이다. 림프절염은 주로 *Mycobacterium avium complex*(MAC), *M. scrofulaceum*에 의해 발병하고 *M. szulgai*, *M. marinum*, *M. ulcerans*, *M. vaccae*은 피부 감염을 일으킨다. MAC, *M. kansasii*, *M. scrofulaceum*, *M. xenopi*등은 파종성 질환의 원인균으로 밝혀졌다. *M. non-chromogenicum*, *M. kansasii*는 관절 및 점액낭 감염을 일으키고 *M. paratuberculosis*는 크론병을 일으킨다. 오랜 세월 인간을 괴롭혀온 결핵균은 예방 및 치료에 대한 연구가 많이 시행되어 발병률이 낮아진 반면, 비결핵 마이코박테리아는 다양한 국가의

도시, 농촌 지역에서 새로운 종이 발견되며 점차 증가하고 있는 추세이다. 근래에는 결핵균 발병률과 비슷한 수준에 이르고 있다. NTM균은 100여종이 발견되었는데 그 중에서 우리나라에서 가장 흔하게 나타나는 균은 *Mycobacterium avium complex*로 대략 60~80%를 차지한다. *M. abscessus*균은 외국에서는 드물게 발견되지만 한국에서는 두 번째로 흔한 NTM균이다. 반면 미국과 일본에서 두 번째로 흔한 *M. kansasii* 폐질환은 국내에서는 상대적으로 발생 빈도가 낮다. 결핵균과 비결핵균 감염을 확인하는 방법은 항산균도말검사, 배양검사, 핵산증폭검사가 있다. 하지만 진단 기간이 수개월에서 수년이 걸리기도 하고 결핵균과 비결핵 마이코박테리아균을 오진하는 경우가 많기 때문에 치료와 진단에 어려움이 있다. 이에 생물정보학적 기술을 이용한 동정이 필요한 실정이다. 최근에는 마커 유전자들을 이용한 종 동정이 다양한 연구에 활용되고 있다. 이는 단일 마커 유전자에 의한 동정보다 정확하고 견고한 것으로 밝혀진 바 있다. 이에 본 연구에서는 비결핵 마이코박테리아 종과 마커 유전자에 대한 서열 정보를 검색하고 마커 유전자 분석에 활용할 수 있는 웹 서버 기반의 데이터베이스를 구축하고자 한다. 이를 이용한 다양한 마커 유전자 서열 조합은 보다 정확한 종 동정으로 이어질 것으로 기대된다. 본 연구의 데이터베이스는 44종의 비결핵 마이코박테리아균 종과 5개의 마커유전자인 16S rRNA, hsp65, rpoB, sodA, tuf에 대한 서열 정보를 효율적으로 검색하고 활용할 수 있도록 하였다. 16S rRNA와 hsp65는 NTM의 가장 대표적인 유전자 마커로 16S rRNA는 모든 원핵 생물이 갖고 있는 유전자이고, hsp65는 생물의 면역력을 발생시키는 마이코박테리아균의 항원 단백질이다. rpoB는 박테리아의 RNA polymerase의 β -subunit 유전정보를 가진 마커 유전자이고, Superoxide dismutase(SOD)는 세포에 해로운 독성을 제거해주는 효소이다. tuf 유전자는 elongation factor EF-Tu의 발현을 담당하는데 그람음성

세균(gram negative bacteria) 종들에게서 *tuf* 유전자의 중복이 발견됨에 따라 박테리아 유전자 분석의 마커 유전자로 이용되고 있다. 기존 연구에서 밝혀진 44종의 비결핵 마이코박테리아 종에 대한 정보를 GenBank의 ftp gbbct 파일에서 JAVA 언어로 추출하여 MySQL 데이터베이스와 fasta 형식의 파일로 저장하였다. 이러한 데이터를 HTML, javascript, JSP 프로그래밍 언어로 가공하여 apache-tomcat 웹 서버와 연동시켰다. 또한 ClustalX, MEGA, SequenceMatrix, WWWBlast 프로그램을 활용하여 상동성 분석, 서열 정렬 및 조합을 통한 계통수 분석을 수행하였다. 연구 결과 16S rRNA-hsp65-rpoB-sodA-tuf의 5개 마커 유전자 조합에 의한 동정법이 16S rRNA-hsp65-rpoB-sodA의 4개 마커 유전자 조합의 경우보다 p-distance 수치가 높고, bootstrap이 50이상인 node수가 많아 더욱 분별력 있고 견고한 방법이 될 수 있음을 확인하였다. 이에 본 데이터베이스와 같은 유전자 서열 정보 시스템을 이용한 다양한 마커 유전자 조합을 통해 보다 정확한 비결핵 마이코박테리아균 동정이 가능함을 증명하였다. 이를 활용한 균 동정은 비결핵 마이코박테리아뿐만 아니라 다양한 균을 대상으로 한 연구로 이어져 해당 균에 대한 예방과 치료법의 개발로 확대될 것으로 기대된다. 이는 인간의 건강 증진을 최우선시 하는 보건학적 측면에서 의미 있는 연구가 될 수 있을 것이다.

.....

주요어 : 비결핵 마이코박테리아, NTM, 생명정보학, 데이터베이스,
계통 분류, 마커 유전자, 보건학

학 번 : 2011-23864

목 차

국문 초록	i
목 차	v
표 목 차	vii
그림 목차	viii

제 1 장. 서 론

1.1 연구의 배경	1
1.2 연구의 필요성	15
1.3 연구의 목적	17

제 2 장. 연구 방법

2.1 데이터 수집 및 가공	23
2.2 시스템 개발 환경 및 데이터베이스 구현	25
2.3 서열 정렬 및 분석 프로그램	28

제 3 장. 연구 결과

3.1 NTM 종과 마커 유전자에 대한 검색 시스템 구축	33
3.2 NTMDB에 의한 마커 유전자 조합 계통분류 분석	36
3.3 WWWblast 분석 결과	42

제 4 장. 고 찰

4.1 고 찰	73
4.2 연구의 한계점	77

제 5 장. 결론 및 총론

5.1 결 론	79
5.2 총 론	86

참고문헌	93
------------	----

Abstract	98
----------------	----

표 목 차

Table 1.1 The Number of NTM isolates in Korea	20
Table 1.2 Factors for NTM disease	21
Table 1.3 NTM species ordered by the location	21
Table 2.1 Marker gene list and database table structure	32
Table 2.2 Analysis programs	32
Table 3.1 Marker gene data status	70
Table 3.2 NTM species data status	70
Table 3.3 The results of marker gene analysis	72
Table 3.4 The result values of WWWblast for <i>M. peregrinum</i>	72

그림 목 차

Figure 1.1 Annual number and rate of isolates of NTM...	19
Figure 2-1 “gbbct.seq” file format	30
Figure 2-2 Database format and Fasta format	31
Figure 3.1 The home page of NTMDB	46
Figure 3.2 “NTM Gene Search” page	47
Figure 3.3 the result of NTM gene searching	48
Figure 3.4 The function of “Marker-sequence Searching” page	49
Figure 3.5 the result of searching other gene, dnaJ	49
Figure 3.6 The function of “NTM Sequence Fasta Files” page	50
Figure 3.7 The function of “Blast & Analysis Programs” pages	50
Figure 3.8 Phylogenetic tree (NJ) of NTM using 16S rRNA gene	51
Figure 3.9 Phylogenetic tree (NJ) of NTM using hsp65 gene	52
Figure 3.10 Phylogenetic tree (NJ) of NTM using rpoB gene	53
Figure 3.11 Phylogenetic tree (NJ) of NTM using sodA gene	54
Figure 3.12 Phylogenetic tree (NJ) of NTM using tuf gene	55
Figure 3.13 Phylogenetic tree (NJ) of NTM using 16S rRNA,...	56
Figure 3.14 Phylogenetic tree (NJ) of NTM using 16S rRNA,...	57
Figure 3.15 P-distance Pairwise matrix using marker gene...	58
Figure 3.16 WWWblast database construction for 44 NTM species	
5 marker genes and blastn and tblastx execution	59
Figure 3.17 The result of blastn for M. peregrinum...	60
Figure 3.18 The result of 16SrRNA blastn	61
Figure 3.19 The result of 16SrRNA tblastx	62

Figure 3.20 The result of hsp65 blastn	63
Figure 3.21 The result of hsp65 tblastx	64
Figure 3.22 The result of rpoB blastn	65
Figure 3.23 The result of rpoB tblastx	66
Figure 3.24 The result of sodA blastn	67
Figure 3.25 The result of sodA tblastx	67
Figure 3.26 The result of tuf blastn	68
Figure 3.27 The result of tuf tblastx	69

제 1 장. 서 론

1.1 연구의 배경

1.1.1 비결핵 마이코박테리아의 정의 및 발병 현황

비결핵 마이코박테리아(non-tuberculous mycobacteria, NTM)는 결핵균(*Mycobacterium tuberculosis*)과 나병균(*M. leprae*)을 제외한 마이코박테리아를 의미한다(Koh et al., 2008). Non-tuberculous mycobacteria, 즉 NTM은 비결핵 항산균(mycobacteria other than tuberculosis, MOTT) 혹은 비정형 항산균(atypical mycobacteria)으로도 불리고 170여종이 넘는 균종을 포함한다. NTM에 비정형, 비결핵이라는 명칭이 붙은 이유도 발견 초기에는 희귀성 결핵균으로 간주되었지만 특정 환경에서만 발견되는 결핵균과 달리 일반 환경에서도 발견된다는 특성에 의해 구분되었다(Primm et al., 2004). AIDS 출현 이전에는 호흡기, 경부 림프선, 피부 등에 주로 발병하였으며, 이 중 90% 이상이 호흡기 질환에 해당하였다(Jo et al., 2010). 호흡기 질환의 경우 60대 남성에게 주로 발병되었고 대부분의 환자들은 폐 상태가 본래 양호하지 않거나 먼지가 많은 농가에서 일하는 경우가 많았다. 대표적인 호흡기 마이코박테리아균으로 *Mycobacterium kansasii*, *Mycobacterium avium*, *Mycobacterium intracellulare* 등이 있으며, 호흡기 질환뿐만 아니라 특히, *Mycobacterium scrofulaceum*의 경우 어린이를 대상으로 한 경부 림프선염의 주요 원인균으로 알려져 있다.

NTM은 1982년에 처음으로 AIDS 환자에게서 검출되었고 그

후로 미국과 유럽의 AIDS환자의 25~50%가 NTM에 감염된 것으로 보고되면서 급격히 증가하였다(Primm et al., 2004). 또한 미국에서는 검사실에서 배양되는 마이코박테리아 중 NTM이 결핵균보다 분리 빈도가 높아졌다(Koh et al., 2008). Figure 1.1을 통해 대만의 경우도 NTM균이 과거에 비해 증가하여 결핵균과 비슷한 수준으로 발견된다는 것을 확인 할 수 있다(Lai et al., 2010). 보통의 NTM 질병은 호흡기 질환이 주를 이루는 반면 AIDS와 같이 면역력 저하로 인한 NTM 질병은 파종성 질환을 나타내며(Primm et al., 2004), 미국에서는 AIDS 환자의 파종성 질환이 10만 명 중 138명, 스위스의 경우는 10만 명 중 330명 정도로 일반 인구의 300배에 해당하는 수치로 나타난다(Alcaide et al., 1997). 환자의 나이, 상태에 따라 다르지만 *M. xenopi*의 호흡기 질환의 경우 사망률은 69%에 이를 정도로 매우 위험적이다(Andrejak et al., 2010).

한국 NTM 환자들을 대상으로 한 연구에서는 Table 1.1과 같이 1992년부터 2002년까지 한국 결핵 연구원에서 동정한 NTM균종이 증가하고 있음을 알 수 있다. 같은 기간에 결핵균 유병률이 1.8%에서 0.5%로 감소한 것을 볼 때 비결핵 박테리아균과 상반된 추세이다. 다른 연구에서는 1990년대 이후 NTM이 분리되는 비율이 증가하여, 일부 병원에서는 임상 검체에서 분리된 마이코박테리아 중 NTM이 20~30% 차지한다고 보고된 바 있다(Koh et al., 2008). 1997년부터 2002년까지 *M. kansasii* 종에 양성 반응을 보인 150명의 호흡기 질환 환자 중 서울 76명, 울산 18명, 부산 16명 인천 16명 이외에 지방 지역에서 4명이 발생하는 등 대부분의 환자들이 대도시에서 발생한 것으로 보고되었다(Yim et al., 2005). 이는 과거에 토양과 밀접한 농가에서 일하던 인간들의 발병률이 높았던 사실과 상반되는 자

료로 NTM이 인간 생활 모든 영역에 밀접한 관련이 있는 병원균임을 시사한다. 환자들의 증상으로는 기침, 호흡곤란, 객혈, 피로 등이 있었고, 가슴 방사선 사진에 공동 장애(cavitary lesion) 현상을 나타내기도 하였다. 절반 이상의 환부가 우상폐정맥에 나타났다. 우리나라에서 가장 흔한 NTM균은 *Mycobacterium avium-intracellulare* complex(MAC)로 50~60%를 차지하고, *M. avium*보다 *M. intracellulare*가 흔하다. 외국에서 상대적으로 드문 *M. abscessus*는 국내에서는 20~30%를 차지하며 두 번째로 흔한 NTM균이다. 반면 미국과 일본에서 두 번째로 흔한 *M. kansasii* 폐질환은 국내에서는 상대적으로 발생 빈도가 드물게 보고되고 있다(Koh et al., 2008).

과거에 결핵균이 인류의 역사 속에서 오랜 세월 동안 생존을 위협했지만, 최근에는 NTM균이 AIDS의 확산을 계기로 인류의 건강에 새로운 위험 요소로 자리 잡았다. 각 나라별 NTM 연구를 통해 알 수 있듯이 결핵균과 구분이 쉽지 않은 NTM균은 전 세계적으로 인간의 거주지와 밀접한 강과 토양 등의 환경에 존재하며, 호흡기를 포함한 다양한 질환을 발생시키며 인류의 건강을 위협하고 있다. 새로운 종들이 발견되면서 NTM균이 인류의 건강에 미치는 영향은 더욱 커지고 있고 NTM균에 대한 다양한 연구들이 이루어지고 있다. 이와 같이 NTM균의 위험성은 인간의 건강 증진을 위한 보건학적인 측면에서 NTM 종에 대한 연구의 필요성을 증가시켰고, 이를 위해 종 분류와 질환에 대해 자세히 알아볼 필요가 있다.

1.1.2 비결핵 마이코박테리아 질병 종류

AIDS가 번창하던 시기 이전부터 NTM에 의한 호흡기 질환에서 대표적인 원인균은 MAC이었다. MAC는 만성적 기관지염, 기관지 확장증 등에서 발견된다. *M. kansasii* 역시 오래 전부터 호흡기 질환의 중요한 원인균이었고, *M. scrofulaceum*은 호흡기 국소 감염과 관련 있다. 병원에서 사용하는 온수에 의한 호흡기 환자 발생과 관련된 것으로 보고된 바 있는 *M. xenopi*는 45℃에 최적으로 발육하는 박테리아로 여러 호흡기 기저 질환과 관련 있다. 그밖에 *M. simiae*, *M. habana*, *M. szulgai*, *M. fortuitum*, *M. vaccae*, *M. malmoense* 등이 있다 (Katoch, 2004). MAC의 경우 완속발육균, 비발색균으로, high-pressure liquid chromatography(HPLC)와 gas-liquid chromatography 분석법으로 동정된다. 배양적, 생화학적 시험으로 *M. intracellulare*와 구분되기도 하지만 *M. avium*과 *M. intracellulare*의 중간 형태의 균주도 존재한다 (Falkinham, 1996). 오늘날에는 rRNA 유전자 서열을 이용한 DNA probe을 이용해서 MAC를 동정하기도 한다. 최적 배양 환경은 pH 5~5.5로 산성을 띤 환경에서 잘 서식하고 이러한 특징 때문에 병원의 온수 공급 시스템에 집중 서식할 수 있다. 또한 MAC의 생존력은 수도 파이프의 아연 도금 처리된 수도 파이프와 관계되었다는 주장도 있다. 호흡기 질환을 일으키는 MAC는 주로 물로부터 미세 입자 형태로 사람의 폐포로 침투한다. 진폐증, 기관지 확장증, 흡연, 음주 이 외에 다른 마이코박테리아균 질병 등이 위험요인으로 알려져 있다. MAC는 다양한 항생제에 대한 면역력과 다수의 종에 의한 감염을 일으켜 효율적인 치료법을 찾기 어렵다. 그래서 식이요법을 병행한 복합적인 항생제 투여가 필요하며, 여러 연구에서 공통적으

로 macrolides와 rifabutin 혼용이 효과적인 것으로 보고되었다(Falkinham, 1996).

림프절염 또한 MAC에 의해 발병하고 서방 국가와 인도에서는 *M. scrofulaceum*이 발병을 한다고 보고된다. *M. szulgai*와 *M. interjectum*도 림프절염 감염을 일으킨다. *M. szulgai*는 1972년에 처음 보고되었고, 일본, 웨일스, 브리티시 콜롬비아, 미국 등의 환자에게서 발견되었다. *M. szulgai*는 37℃에서는 암발색균이지만 25℃에서는 광발색균인 독특한 마이코박테리아균이다. 다른 암발색균과 다르게 Tween 80이라는 계면 활성제를 가수 분해 한다. 16S rRNA 유전자 서열 분석에 의하면 *M. malmoense*와 밀접한 관련이 있다(Rogall et al., 1990). 위험 요인은 MAC와 비슷하고, isoniazid, rifampin, ethambutol, streptomycin이 치료에 주로 쓰인다. *M. scrofulaceum*은 과거에 어린이들의 경부 림프선염과 관련이 있었고 치료는 보통 절제술이 선호된다. 최근에는 *M. scrofulaceum*보다 MAC에 의한 경부 림프선염이 많이 보고되고 있다. MAC와 유전적, 배양적, 생화학적 관련성이 많은 *M. scrofulaceum*은 완속발육균이고 Tween 80을 가수분해하지 못한다. *M. scrofulaceum*은 MAC와 같이 플라스미드를 갖고 있지만 별개의 종이다. *M. scrofulaceum*의 발병률이 낮아지는 만큼, 주변 환경에서의 개체수도 줄어들고 있다. 일부 연구에 따르면 가공 전 우유와 식물 조직에서 *M. scrofulaceum*이 발견되기도 한다. 치료는 기본적으로 MAC와 비슷한 항생제를 사용하고, 그 외에 clarithromycin이 치료에 이용되고 있다(Falkinham, 1996).

피부 감염은 *M. szulgai*, *M. marinum*, *M. ulcerans*, *M. vaccae*에 의해 발병한다. *M. marinum*은 수영장이나 수족관에서 서식하는 원인균으로 온몸에 구진성 병소를 일으킨다(Katoch, 2004). *M. marinum*은

30℃ 이하의 국한된 곳에 서식하기 때문에 발병률도 산발적이다. *M. marinum*은 발육속도는 완속과 신속의 중간 단계이지만 광발색균에 속하고 catalase 와 nitrate reductase activities가 없다. 일부는 Tween 80 을 가수 분해 한다. 위험 요인으로서는 면역력 저하된 사람이 *M. marinum*이 서식하는 물에 노출되거나 찰과상을 입는 경우이다. 또한 수산 양식업에 종사하는 사람이 찰과상을 입고 *M. marinum*이 서식하는 물이나 어류에 노출되면 감염되기도 한다(Falkinham, 1996). 치료법은 minocycline 또는 doxycycline, trimethoprim-sulfamethoxazole, 또는 rifampin과 ethambutol을 이용한 식이요법이다. 다른 방법으로 rifampin, ethambutol, amikacin을 이용하는 방법이 있다. *M. marinum*은 isoniazid와 pyrazinamide에 저항성이 있고, streptomycin에 일정한 감수성이 있다. *M. ulcerans*는 피부를 통해 진피로 침투하여 감염을 일으킨다. *M. ulcerans*에 의한 피부 궤양은 1928년 호주에서 처음으로 보고되었으나 대부분은 나일 강 근처의 아프리카나 우간다에서 발생하였다. *M. ulcerans* 질환의 공통된 점은 열대 지역의 습지나 강 주변에서 발생한다는 것이다. *M. ulcerans*는 25℃~37℃에서 완속발육한다. 위험요인은 강 주변에서의 농경 활동이고, 긴 바지를 입거나 BCG 백신 투여 등으로 예방할 수 있다. 치료법은 p-aminosalicylic acid와 ethambutol로 가능하고 isoniazid와 rifampin에는 저항성이 있다 (Falkinham, 1996).

파종성 질환의 원인균으로는 *MAC*, *M. kansasii*, *M. scrofulaceum*, *M. xenopi*, *M. simiae*, *M. fortuitum*-*M. chelonae* complex, *M. genavense*, *M. haemophilum* 등이 있고 그 외에 *M. celatum*, *M. conspicuum*, *M. malmoense* 등이 있다. 파종성 질환은 AIDS와 같은 면역력 결핍 환자에서 많이 나타나고 있다. *M. simiae*는 1965년 마이코

박테리아균에 감염된 원숭이에게서 처음 발견되었고, 여러 연구에서 AIDS 환자의 *M. simiae*균에 대해 보고한 바 있다(Falkinham, 1996). *M. simiae*는 광발색균으로 37℃에서 pH 5.5~7.5에서 가장 잘 배양된다. 다른 비결핵 마이코박테리아균과 달리 niacin을 생성하고 urease를 지니고 있다(Katoch, 2004). *M. simiae*는 독특하게 16S rRNA 서열이 완속발육균과 신속발육균 특징 모두를 지니고 있고, Tween 80 가수분해, 색소 생성의 특징으로 MAC와 *M. scrofulaceum*과 종종 잘못 분류되기도 한다. *M. simiae*는 감염된 원숭이와 같은 우리 안에 생활한 26%의 건강한 원숭이들이 감염됐다는 연구가 보고됨에 따라 동물 간에 전염성이 알려졌다. 구체적인 전염 메커니즘은 알려지지 않았지만 동물의 배설물이 구강을 통해 전염될 가능성이 높다(Falkinham, 1996). 다른 비결핵 마이코박테리아균처럼 물에서 동정되기도 한다. MAC와 같은 약물 치료법이 알려져 있으나 효과는 불확실하다(Falkinham, 1996).

*M. xenopi*는 1990년 프라하에서 객담 환자들에게서 발견되었다. 호주, 일본, 짐바브웨 등에서 드물게 발견되고 캐나다, 웨일즈, 북서 연안 유럽에서는 의료용품에서 발견된다. *M. xenopi*의 가장 큰 특징은 발육 최적 온도가 45℃이고 불규칙한 막대 형태를 하고 있다. Arylsulfatase는 발육온도와 색소생성이 비슷한 *M. avium*과 구분되는 *M. xenopi*의 특징이다. 그밖에 16S rRNA와 HPLC로 MAC와 구별이 가능하다. *M. xenopi*는 역학적으로 상수도에 주로 서식할 것이라는 추측되고 실제로 수돗물이나 샤워 꼭지에서 발견되기도 한다. 그 외에 물탱크, 식수 처리 시설, 상수도 시설에서 발견되지 않는 것은 *M. xenopi*가 28℃ 이하에서 발육할 수 없기 때문일 수 있고 이것은 MAC와 구분되는 점이기도 하다. *M. xenopi* 감염의 위험요인은 AIDS,

폐질환, 흡연, 담배, 만성 간 질환, 심혈관 질환, 위절제술 등이 있다. *M. xenopi*의 약물치료는 성공률이 떨어지지만, rifampin, streptomycin, isoniazid, ethambutol 등의 조합으로 효과를 보기도 한다(Falkinham, 1996). *M. szulgai*, *M. fortuitum*, *M. non-chromogenicum*, *M. kansasii*는 관절 및 점액낭 감염을 일으키고 *M. paratuberculosis*는 크론병의 원인이 된다(Katoch, 2004). AIDS 감염 확산이 MAC의 감염을 크게 증가시켰지만, AIDS 확산 이전에 *M. kansasii* 감염이 더 많았다. 텍사스에서는 1967년부터 1976년까지 *M. kansasii* 감염이 크게 증가하여 MAC의 감염 건수의 2배로 보고된 바 있다(Falkinham, 1996). *M. kansasii*는 32℃~42℃에 완속발육 하는 광발색균이다. 분리균은 catalase와 nitrate reductase를 생성하고 Tween 80을 가수분해한다. *M. kansasii*는 물과 흙에서 발견되고 물에서는 12개월까지 살 수 있지만 흙에서는 긴 시간 동안 살 수 없다. 위험 요인은 호흡기 질환, 암, 알코올 중독이 있고, 낭포성 섬유증 환자의 감염은 유전병 또한 위험 요인이 될 수 있음을 시사한다. 약물 치료는 다른 비결핵 마이코박테리아 치료와 마찬가지로 여러 약물을 조합을 맞추어 같이 복용함으로써 치료한다. rifampin- streptomycin, isoniazid, ethambutol, 또는 rifampin, ethambutol, amikacin, 또는 rifampin, isoniazid, ethambutol, 추가적으로 streptomycin을 18개월 간 투여하기도 한다(Falkinham, 1996). 크론병은 위장관에 만성 염증을 일으키는 질병이다. 크론병을 일으키는 *M. paratuberculosis*는 가축의 장에서 설사와 소모성 질환을 일으키는 일부 결핵증의 원인균이기도 하다. *M. paratuberculosis*는 MAC와 비슷하지만 마이코박틴 의존도와 IS900 유전자로 구별된다. *M. paratuberculosis*는 주로 가축들 간에 전염으로 확산되지만, 전염된 가축을 통해 인간이 전염되기도 한다. 위험 요인으로서는 전염된 가축의

우유나 물, 토양 등이 있다(Falkinham, 1996).

그밖에 Table 1.2와 같이 NTM 질병은 인간의 면역력, 생태시스템, 여행, 공중 보건, 전쟁, 기근 등의 다양한 위험요인이 있다. NTM 질환의 발병은 토양, 물과 같은 지리적인 요인뿐만 아니라 고령화, 여행, 전쟁, 기근과 같은 인류의 문화적 요인도 포함하고 있어 NTM균의 발병률 증가와 질환의 심각성을 고려해 볼 때 확산을 막기 위한 연구의 진행이 필요함을 알 수 있다. 이를 위해서 발병 역학적인 연구와 함께 근본적으로 NTM균에 대한 종과 유전자에 대한 정보를 파악하고 분석할 필요가 있다. 이는 확산을 방지할 뿐 아니라 발병을 예방하고 새롭게 발견된 종에 대한 분석도 수월하게 진행하기 위함이다. Table 1.3은 NTM 종별로 발견되는 나라를 표로 나타낸 것이다. 많은 나라와 다양한 NTM균의 발생은 전 세계 모든 국가의 보건 문제와 즉결되었다고 해도 과언이 아님을 보여주고 있다. 모든 균은 진화하고 더 위험한 균이 새롭게 발견되기 마련이기 때문에 NTM에 대한 연구는 모든 관련 연구 기관들이 필요성을 알고 관심을 가져야만 한다. 이와 같이 NTM균은 지리적 문화적 요인으로 전 세계 대부분의 나라에서 다양하게 발견되고 있다. NTM균 종류와 질환의 종류에 따라 진단 및 치료법이 다르기 때문에 균 동정이 매우 중요하다. 하지만 지금까지는 균의 특성상 동정이 매우 불분명하고 시간도 매우 오래 걸렸기 때문에 관련 질병 치료를 위한 빠른 분류와 동정에 대한 연구가 필요한 실정이다.

1.1.3 NTM 마커 유전자에 대한 연구 현황

NTM의 분류는 마이코박테리아 전통적인 분류법과 같이 Runyon classification에 의해 이루어졌다. 증식 속도에 따라 10일 전후로 신속발육군(rapid grower)과 완속발육군(slow grower)으로 구분하고 완속발육군은 색소생성에 따라 비발색균(nonchromogen), 광발색균(photochromogen), 암발색균(scotochromogen)으로 분류한다. NTM균은 도말 검사나 배양검사로 결핵균과 구분이 어렵지만 PCR 검사로 구분 가능하다(Lee et al., 2007). NTM 연구를 시작하던 1960 ~ 70년대에는 주로 색소를 이용한 간단한 수학적 모델을 활용하거나 Carotenoid 와 같은 균의 색소를 이용하여 스펙트럼 분석하거나 생화학적 검사를 활용한 분류가 많이 이루어졌다(Lee et al., 2007; Tsukamura, 1968). 당시에는 균을 형태학적 분류법과 화학물질에 반응하는 양상에 따른 분류 이외에는 기술적으로 한계가 있었기 때문에 이러한 방법에 의존했고 각각의 균주를 관찰하고 분리해야 하기 때문에 시간과 인력이 많이 필요했다. 하지만 1980년대 AIDS 확산과 함께 NTM 질병이 크게 확산되었고 1990년대 전 세계적으로 활발하게 진행된 유전자 서열 분석의 발전에 힘입어 NTM 종의 분류법도 마커 유전자 서열 정보를 이용하기 시작하였다.

rRNA는 단백질을 합성하는 소기관인 리보솜을 구성하는 RNA로서 세포 내 RNA의 대부분을 차지한다. rRNA는 원핵 생물과 진핵 생물에 따라 사이즈가 다르다. 이 중 16S rRNA는 모든 원핵 생물이 갖고 있는 유전자로 박테리아 분류에 많이 쓰이는 대표적인 마커 유전자이다. 1990년대에 마이코박테리아 속(genus)을 16S rRNA 서열을 이용하여 계통분류를 분석한 연구가 나왔다(Rogall et al., 1990;

Springer et al., 1995). 1996년 NTM 감염에 대한 역학 연구에서는 16S rRNA 서열 분석에 의해 NTM의 발육속도에 의한 분류가 가능하다고 하였다(Falkinham, 1996). 국내에서도 16S rRNA 서열을 이용하여 하수 처리시설의 세균 군집구조 분석(Kim et al., 2010), 해삼물의 장내 세균의 다양성(Noh et al., 2013), 남조세균 혼들 말목 해양 균주(Cheon et al., 2012) 등 다양한 박테리아균 분석 연구를 시행하였다.

그 후 2003년 마이코박테리아균의 분류법에 대한 연구에서는 마이코박테리아균을 대상으로 1990년대부터 발전하기 시작한 유전자 분석법의 활용도를 확인하였다. Runyon classification에 의해 색소와 성장속도로 구분된 마이코박테리아 40여종을 가장 대표적인 유전자 마커인 16S rRNA와 hsp65를 이용하여 분석하였다(Tortoli, 2003). 마커 유전자 서열 분석법과 기존의 다양한 화학적 분류법은 NTM을 포함한 마이코박테리아균 분류에 각각의 역할이 있고 어느 한 가지 방법만으로는 정확한 결과를 기대할 수 없다고 결론짓고 있다. 하지만 hsp65 유전자를 이용한 방법은 서열 분석이 아니라 전기영동(agarose gel electrophoresis)을 이용한 띠 형성을 기준으로 한 화학적 분류법에 가까웠다. 유전자 서열 분석에 사용된 것은 16S rRNA 유전자뿐 이었고 연구 결론에서 16S rRNA를 이용한 방법은 어느 정도 유효한 결과를 나타내지만 분석에 사용하는 유전자 서열 구간에 따라 계통 분류도 결과가 달리 나오는 것이 지적되었다. 또한 일부 마이코박테리아 종들은 16S rRNA에 의한 구분도 명확하지 않았고, 이로써 복수의 다양한 유전자 조합에 의한 분석의 필요성이 제기되었다.

16S rRNA 이외에 다양한 마커 유전자가 발견됨에 따라 이들을 이용한 마이코박테리아균 분석 연구들이 진행되었다. 그 중 heat

shock protein 65(hsp65)는 16S rRNA 다음으로 마이코박테리아균의 동정에 많이 쓰이는 마커 유전자이다. Heat shock protein은 열에 반응하는 단백질 그룹을 말하고 다른 단백질이 활성화 하기 위한 접힘(folding)에 관여한다. 특히 hsp65는 생물의 면역력을 발생시키는 마이코박테리아균의 항원 단백질이다. Hsp65를 이용한 유전자 서열 분석 연구에 의하면 다른 분석법과 80% 이상 일치하였고 생화학적 분석법에 비해 신속하고 경비를 절감할 수 있는 분석법이였다. 또한 hsp65 서열 분석법은 16S rRNA보다 종 구분이 확실하고 마이코박테리아균의 계통 분류도와 종 동정에 유용하였다(Kim et al., 2005; McNabb et al., 2004). RpoB는 박테리아의 RNA polymerase의 β -subunit 유전정보를 가진 마커 유전자이다. Rifampin은 가장 대표적인 결핵 치료 항생제로서 원핵생물의 RNA polymerase와 반응하여 전사(transcription)를 방해하여 세포사(cell death)를 일으킨다. rpoB의 변형은 rifampin과 RNA polymerase의 결합력을 감소시켜 rifampin에 대한 내성을 갖게 한다(Hirano et al., 1999). rpoB는 16S rRNA로는 동정이 어려운 *M. kansasii*와 *M. gastri* 같이 분류학적으로 밀접한 마이코박테리아균 종들의 동정에 효과적이고 16S rRNA를 보완할 수 있는 마커 유전자이다(Case et al., 2007; Kim et al., 1999).

Superoxide dismutase(SOD)는 생물의 일반적인 물질대사 후 생성된 초과산화이온($O_2^{\cdot -}$)을 산소(O_2)와 과산화수소(H_2O_2)로 분해하여 세포에 해로운 독성을 제거해주는 효소이다. SodA는 SOD의 일종으로 효소의 활성 부위가 망간(manganese)과 결합하기 때문에 manganese SOD, 또는 sod2로도 표기한다(Melov et al., 2001; McCord et al., 1969). SodA는 주로 16S rRNA, hsp65, rpoB와 같은 다른 마커 유전자와 조합하여 분석에 사용된다(Devulder et al., 2005). recA 단백질은

상동 DNA 재조합과 손상된 DNA의 변형을 보수하는 유전자이다. recA는 상동 DNA의 binding, pairing, exchanging 뿐만 아니라 ATP 가수 분해, 단백질의 cleavage 역할도 담당하고 있다. recA 단백질의 중요한 역할로 인해 일정한 진화적 거리가 있는 박테리아 간의 분류 계통분석에 쓰이는 마커 유전자로 제기 되고 있다(Torriani et al., 2001). 실제로 NTM균을 포함한 마이코박테리아균의 동정 분석에 recA 유전자를 이용하여 연구가 이루어지기도 하였다(Blackwood et al., 2000).

tuf 유전자는 아미노아실 tRNA(aminoacyl-tRNA)가 codon recognition site에 결합하여 단백질을 생성하는 elongation factor EF-Tu의 발현을 담당한다(Filer et al., 1980). 그람음성 세균(gram negative bacteria) 종들에게서 tuf 유전자의 중복이 발견됨에 따라 박테리아 유전자 분석의 마커 유전자가 되었다(Sela et al 1989; Filer et al., 1981). 마이코박테리아균의 동정 분석에서도 tuf는 hsp65와 비슷한 분류 계통도를 보이며 임상 분리균의 동정에 정확성을 보였다(Mignard et al., 2007). 그 외에도 tmRNA encoding 유전자(ssrA), dnaJ 유전자, gyrB 유전자가 있으며 소수 종에 대해서는 ESAT-6 유전자 등이 NTM을 포함한 마이코박테리아균의 종 동정에 사용되고 있다(Mignard et al., 2007; Takewaki et al., 1993; Kasai et al., 2000; Singh et al., 2007).

하지만 각 유전자를 이용한 연구들은 조금씩 다른 결과를 보여줬고 어느 하나의 마커 유전자에 대한 연구만을 정설로 여길 수 없다. NTM균 분류에 대해 각 마커 유전자가 실제로 어떤 비중으로 어떤 역할을 하는지 현재까지는 알 수 없다. 지금까지 균을 포함한 생물 유전에 대한 연구는 생물에 대한 완벽한 법칙에 다가가기 위해 유전자를 기반으로 확률과 통계의 개념을 이용하여 진행되었

다. 그러기 위해 관련 연구자가 할 수 있는 일은 좀 더 많은 데이터를 수집하고 하나의 분석 틀에 적용하여 자료와 분석의 변별력과 견고함을 높이는 것이다. 최근에는 다양한 마커 유전자를 같이 비교하거나 조합하여 한 가지 유전자만을 이용한 분석의 한계점을 극복하고, 보다 정확한 NTM을 포함한 마이코박테리아균 동정 분석법에 대한 연구가 이루어지고 있다. NTM균을 대상으로 16S rRNA, hsp65, rpoB 유전자를 각각 이용하여 비교하거나 16S rRNA-hsp65-rpoB-sodA 조합을 이용한 multigene approach를 통하여 분석하기도 한다 (Devulder et al., 2005; Mignard et al., 2007).

1.2. 연구의 필요성

NTM은 자연계에 감염 보유 숙주가 없는 결핵균과 달리 물과 토양과 같은 주변 환경에 분포하고 있고 사람과 사람 사이에서의 전염이 없다. NTM은 균종에 따라 질환을 일으키는 발병력(virulence)이 다르다. *M. kansasii*, *M. abscessus*는 발병력이 상대적으로 크고, *M. fortuitum*은 상대적으로 낮다. 따라서 호흡기 검체에서 NTM이 분리되었다고 해도 이것이 NTM 폐질환의 증거라고 할 수 없다. 임상 검체에서 NTM이 분리되었을 때 임상적 의의가 있는 감염증을 나타내는 비율은 진단기준과 국가에 따라 차이를 보인다. 미국, 캐나다, 서유럽에서는 NTM이 발견된 환자 중 약 40~50%가 NTM 질환을 가진 것으로 보고되었고, 국내 연구에서는 객담에서 NTM이 분리된 환자 중 10~25%가 NTM 폐질환을 가지고 있는 것으로 알려졌다. 검체에서 분리된 NTM은 임상적, 방사선학적, 미생물학적 기준의 진단이 필요하다. 미국 흉부 학회(American Thoracic Society)와 미국감염학회(Infectious Diseases Society of America, IDSA)에 의해 NTM 폐질환의 진단 기준이 발표되기도 하였다. NTM 폐질환의 방사선학적 이상소견의 변화속도가 매우 느리기 때문에 검체에서 분리 시 충분한 추적관찰기간이 필요하다. 수개월 간격으로 관찰하고 심지어 수년이 지난 후에 발견되기도 한다(Korea Center for Disease Control & Prevention, 2011). NTM은 결핵균과 비슷한 증상을 보이지만, NTM(Non-tuberculous mycobacteria) 단어가 의미하듯이 마이코박테리아균 속(genus)에서 결핵과 구분되는 별도의 균을 의미하기 때문에 진단과 그에 따른 치료 연구를 위해서는 정확한 균 동정이 필요하다.

AIDS의 확산과 함께 전 세계적으로 NTM 환자가 증가하기 시작하였고 국내에서도 NTM 폐질환으로 진단 치료 받는 환자들이 증가하고 있다. NTM 질환은 인간의 각 기관을 파괴하고 사망을 초래하는 병이고, 수개월에서 수년이 걸리는 진단 기간과 함께 결핵균과 구분이 어렵다는 점에서 인간에게 매우 위협적이다. 2000년대에 인간 게놈 프로젝트의 결실로 유전자 서열 연구에 대한 기대가 매우 커졌고 최근에는 차세대 유전자 서열 분석 기술인 NGS(Next Generation Sequencing) 기법의 발달로 NTM에 대한 분석법도 많은 변화가 생겼다. 기존의 임상적, 방사선적, 생화학적, 형태학적 동정법과 더불어 유전자 서열 분석에 의한 생명정보학적 동정법이 많이 사용되고 있다. 국내에서는 결핵 연구원, 제주대학교, 서울대학교에서 비결핵항산성균에 대한 연구를 하고 있다. Global Health Institute at the Ecole Polytechnique에서 결핵균에 대한 검색 DB를 구축하였지만 아직까지는 NTM에 대한 오픈 소스의 데이터베이스가 없는 실정이다. 공신력 있는 공통된 분석 데이터베이스가 있다면 각 국가 각 지역에서 발생한 NTM 환자에게서 발견된 균주의 동정이 더욱 효율적이고 객관적이며 결과 분석에 일관성을 나타낼 것이다. 보건학적으로 NTM 질병의 빠른 진단 및 치료를 위해 연구자가 임의로 선택한 마커 유전자 조합이 가능한 NTM 데이터베이스가 절실히 필요하다.

1.3. 연구의 목적

본 연구의 목적은 NCBI GenBank의 NTM 종에 대한 유전 정보, 마커 유전자의 서열 정보를 담고 있는 데이터베이스를 구축하여 검색 및 다운로드 기능을 가능하게 하고, 관련 프로그램을 통한 서열 정보 분석과 계통 분류학적 분석을 할 수 있도록 하는 것이다. 데이터베이스에 구축된 NTM 종과 마커 유전자에 대한 정보 검색과 특정 NTM 종이 지니는 유전자와 특정 유전자를 지닌 NTM 종에 대한 검색 역시 가능하게 하여 연구자가 유동적으로 필요한 정보를 검색하고 데이터를 다운로드 받을 수 있도록 한다. 또한 본 데이터베이스 웹 서버에서 clustalX, SequenceMatrix, MEGA 분석 프로그램을 이용할 수 있도록 하여 연구자가 복수의 NTM 마커 유전자를 정렬 및 조합하고, 계통수 분석과 같은 다양한 연구에 활용할 수 있도록 제공하고자 한다. 현재 NTM을 연구하는 기관들은 존재하지만 NTM만의 오픈 소스 데이터베이스가 없기 때문에 본 연구는 각 연구 기관에 산재되어 있는 데이터들을 통합하는 계기가 될 것이며, 이를 통해 진단 기간이 긴 NTM 질환의 치료에 획기적인 발전의 계기를 마련하고자 한다.

이와 같은 NTM 종과 유전자에 특화된 보건학적 데이터베이스 구축은 근래에 발병이 증가하여 인간을 위협하고 있는 NTM 질환의 신속 정확한 동정을 가능하게 함으로써 NTM 관련 질환의 진단과 치료에 기여할 수 있을 것이다. 이 뿐만 아니라 각 나라 각 지역에서 새롭게 발견되는 NTM 종에 대한 정보를 공유함으로써 기존에 알려진 종들과 비교 분석하여 새로운 종의 분류학적인 위치 분석을 가능하게 할 것이다. 또한 이와 같은 방법은 다른 질병 군에

대한 데이터베이스 구축에 대한 모티브로서 기여할 수 있을 것으로 기대된다. 웹 서버 기반의 본 데이터베이스가 NTM 종 분석법에 기여하여 인간의 건강 증진을 위한 보건학적인 수단이 되도록 하는 것이 본 연구의 목적이다.

이러한 연구를 수행하는데 있어 생명정보학적 접근이 필요한 이유는 지금까지 밝혀진 NTM의 종과 유전자의 방대한 생물학적 정보를 분석 및 처리하는데 컴퓨터 연산 기술이 적용되어야 하기 때문이다. 본 연구는 결핵균과 유사하고 종 동정에 오랜 시간이 걸리는 기존의 비결핵 마이코박테리아균 동정을 대신하여 생명정보학적 기법을 활용하여 보다 신속하고 정확한 종 동정법을 연구자에게 제공하고자 한다. 이는 앞으로 개척해 나가야 할 부분이 많은 비결핵 마이코박테리아균을 포함한 다양한 박테리아균 종 분류와 마커 유전자 분석에서의 생명정보학적 기법 활용에 기여할 것이고 이는 생물 종에 대한 모든 연구에 크게 이바지할 것으로 기대된다.

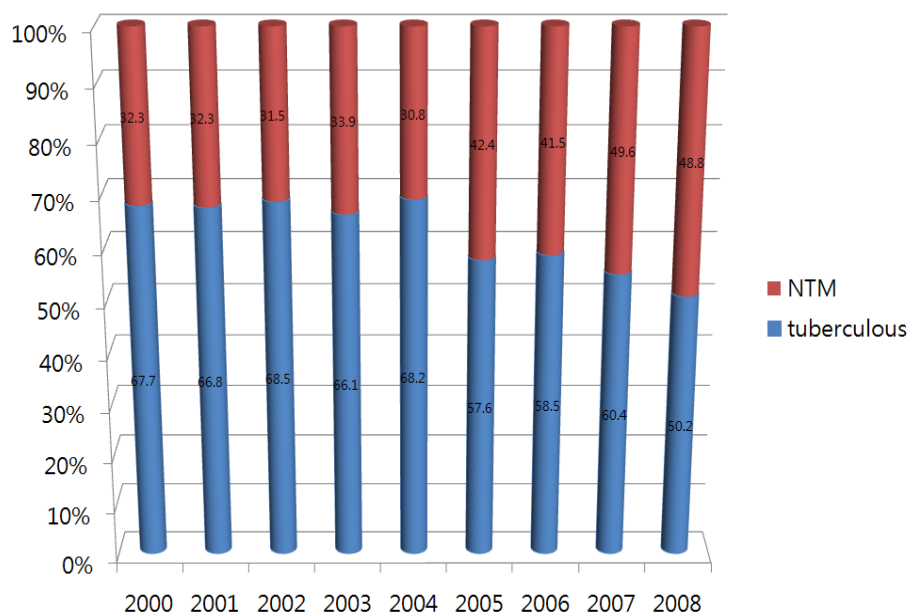


Figure 1.1 Annual number and rate of isolates of NTM, National Taiwan University Hospital. The number of NTM isolates has been increased and reached the level of tuberculosis isolates (Lai et al., 2010).

Table 1.1 The number of NTM isolates in Korea

year organism	19 92	19 93	19 94	19 95	19 96	19 97	19 98	19 99	20 00	20 01	20 02
<i>M. kansasii</i>	1	9	21	18	16	12	22	26	35	47	62
<i>M. avium- intracellulare</i>	245	302	329	287	266	253	551	499	593	693	769
<i>M. fortuitum</i>	49	66	60	41	27	20	47	62	56	92	184
<i>M. chelonae</i> complex	32	34	43	55	23	28	63	92	88	166	279
<i>M. chelonae</i>	n-a	n-a	n-a	n-a	n-a	n-a	n-a	n-a	n-a	8	18
<i>M. abscessus</i>	n-a	n-a	n-a	n-a	n-a	n-a	n-a	n-a	n-a	158	261
<i>M. scrofulaceum</i>	1	3	15	2	3	0	4	0	0	0	2
<i>M. terrae</i> complex	14	25	43	27	8	5	32	25	35	3	52
<i>M. szulgai</i>	3	2	9	5	0	6	6	13	18	31	15
<i>M. gordonae</i>	5	12	21	10	15	9	22	39	18	31	126
<i>M. celatum</i>	0	0	0	0	0	0	12	8	11	7	10
<i>M. marinum</i>	0	0	0	0	0	2	1	0	1	1	0
<i>M. smegmatis</i>	0	0	0	0	0	0	0	0	0	0	0
Unidentifiable	99	250	608	403	558	752	422	400	314	559	63
Total	448	791	1,149	848	916	1,087	1,182	1,164	1,169	1,611	1,737

The number of NTM patients has been increased in Korea (Yim et al., 2005).

Table 1.2 Factors for NTM disease

Factor	Effect on nontuberculous mycobacterial disease
Microbial adaption	Adapt to survival in the presence of antimicro
Changes in human susceptibility	Increase in immunodeficient host
Changing ecosystems	Increased demand for water
changing human demographics	Aging population of increased susceptibility
Land use	Increased opportunity for transmission / new environments for mycobacterial growth
Travel	Worldwide transmission of novel mycobacteria
Breakdown of public health	Inability to monitor mycobacterial disease outbreaks
Poverty	Increase in susceptible hosts, less access to care
War, dislocation and famine	Increase in susceptible hosts, less access to care, greater opportunity for transmission

Most of factors for NTM diseases are environmental (Falkinham, 2003).

Table 1.3 NTM species ordered by the location

Common	Comment	Uncommon	Comment
Pulmonary Disease			
M. abscessus	Worldwide; may be found concomitant with MAC	M. asiaticum	Rarely isolated
M. avium complex	worldwide; most common NTM pathogen in U.S.	M. celatum	Cross-reactivity with TB-DNA probe
M. kansasii	U.S., Europe, South Africa, coal-mining regions	M. fortuitum	Associated with aspiration
M. molmoense	U.K., northern Europe; uncommon in U.S.	M. haemophilum	Rarely isolated
M. xenopi	Europe, Canada; uncommon in U.S.; associated with pseudoinfection	M. scrofulaceum	South Africa; uncommon in U.S.
Lymphadenitis			
M. avium complex	Worldwide; most common NTM pathogen in U.S.	M. abscessus	Rarely isolated
M. molmoense	U.K., northern Europe (especially Scandinavia)	M. genavense	Fastidious species (See Laboratory Procedures)

M. scrofulaceum	Worldwide; previously common, now rarely isolated in U.S.	M. haemophilum	Fastidious species (See Laboratory Procedures)
Disseminated Disease			
M. avium complex	Worldwide; AIDS; most common NTM pathogen in U.S.	M. abscessus	Non-AIDS immunosuppressed
M. chelonae	U.S.; non-AIDS immunosuppressed skin lesions	M. conspicuum	AIDS, non-AIDS immunosuppressed
M. haemophilum	AIDS; U.S., Australia; non-AIDS immunosuppressed	M. fortuitum	Non-AIDS immunosuppressed
M. kansasii	AIDS; U.S., South Africa	M. genavense	AIDS
Skin, Soft Tissue, and Bone Disease			
N. abscessus	Penetrating injury	M. avium complex	Worldwide
M. chelonae	U.S., associated with keratitis and disseminated disease	M. haemophilum	Extremities, cooler body sites
M. fortuitum	Penetrating injury, footbaths	M. immunogenum	Rarely isolated, associated with pseudo-outbreaks
M. morinum	Worldwide, fresh-and saltwater	M. kansasii	Rarely isolated
M. ulcerans	Australia, tropics, Africa, Southeast Asia, not U.S.	M. molmoense	U.K., northern Europe
NTM species exist all over the world and cause several the diseases on the several part of human body (Griffith et al., 2007).			

제 2 장. 연구 방법

2.1 데이터 수집 및 가공

본 연구에서 사용된 데이터는 NCBI의 GenBank ftp 사이트 (<ftp://ftp.ncbi.nih.gov/genbank/>)에서 수집 하였다. ftp는 파일 전송 프로토콜(File Transfer Protocol)로서 서버와 클라이언트 사이의 대용량 파일 전송을 하기 위한 프로토콜이다. NCBI는 지금까지 밝혀진 모든 생물의 대용량 유전 정보를 모두 보유하고 있고 이러한 데이터를 연구자들이 빠르고 쉽게 받을 수 있도록 ftp 방식을 사용하고 있다. 이 중 모든 박테리아의 유전 정보를 담고 있는 142개의 gbbct 파일을 다운 받아 NTM 종에 대한 데이터를 추출하기 위해 JAVA 프로그래밍 언어를 이용하여 데이터를 추출하였다.

gbbct 파일은 박테리아에 대한 유전 정보를 담고 있는 파일이다. 2014년 8월 19일 기준으로 gbbct1.seq~gbbct142.seq까지 총 142개의 파일로 구성되어 있다. Figure 2-1과 같이 gbbct 파일에는 organism, definition, version, reference, gene region, gene 명칭, protein ID 등 종과 유전자에 대한 자세한 정보가 담겨있다. 이를 JAVA 언어를 이용하여 데이터베이스 구축에 필요한 종과 그에 대한 데이터를 추출하였다. 웹 서버에 데이터베이스 형식의 데이터를 입력하기 위해 accession, organism, definition, CDS, gene, product 항목을 추출하여 탭(tab)으로 구분을 하였다. Accession은 서열 고유 번호, organism은 종명, definition은 종과 서열에 대한 구체적인 설명, CDS는 coding DNA

sequence 부분의 좌표, gene과 product는 각각 유전자 이름과 유전자 산물을 나타낸다. 이를 사용자가 검색 가능한 MySQL 데이터베이스에 파일 형식으로 저장하여 사용자가 구체적인 서열 정보를 검색할 수 있도록 하였다. 분석을 위한 종과 유전자의 서열 정보는 분석이 용이하도록 organism, accession, gene, product, DNA sequence 순으로 fasta 형식으로 추출하였다. Fasta 형식은 생명정보학에서 주로 쓰이는 파일 형식으로, 기호 “>” 로 시작하여 gene id, 종 명, 유전자 명 등과 같은 종에 대한 각종 정보가 나열되고 다음 줄에 서열 정보가 나오는 형식이다. Fasta 형식의 데이터는 기호 “>” 사이의 정보를 한 단위로 취급한다. Figure 2-2는 Figure 2-1의 gbbct 파일을 JAVA language를 이용하여 데이터베이스 형식과 fasta 형식으로 추출한 결과이다. 추출한 데이터와 파일들은 웹 프로그래밍 언어인 HTML, JSP, javascript를 이용하여 서버에 연동 하였다. WWWBlast (Altschul et al., 1990)는 command prompt 창에서 관련 명령어로 fasta 파일을 기입하여 분석에 이용하였다.

2.2 시스템 개발 환경 및 데이터베이스 기능

NTM 동정을 위한 마커 유전자 검색 및 조합 시스템 구축을 위해 Intel(R) Core(TM) i5-2537M 1.40GHz, 4Gb RAM, 100Gb HDD 사양의 서버를 구축하였다. Operating System은 Window 7 Home Premium K를 사용했고 데이터베이스 구축을 위해 MySQL 서버 5.0, 웹 서버 프로그램으로 Apache-tomcat 7.0을 설치하였다. 또한 웹 서버와의 연동을 위해 웹 프로그래밍 언어인 HTML, JSP, javascript를 사용하였다. 본 연구는 연구자가 웹 서버를 이용해서 NTM 종과 유전자 검색을 통해 서열 정보를 손쉽게 얻고 연구 목적에 맞게 다룰 수 있는데 목적이 있다. 주어진 분석 프로그램을 이용하여 다양한 마커 유전자 서열을 조합하고 이를 통해 다양한 동정법이 가능하도록 하고자 한다. 이를 위해 데이터베이스 구현은 가장 중심이 되는 연구 과정이다. 본 데이터베이스에 구현된 NTM 종은 결핵 연구원에서 2009년부터 2011년까지 우리나라 임상환자에서 분리한 종들과 제주대학교에서 분리한 종들을 합친 44종을 대상으로 하였다(Kim et al., 1999; 유 외, 2011). 결핵 연구원의 *M. massiliense*와 *M. kumamotense*는 gbbct.seq 파일에 포함되어있지 않아서 대상에서 제외하였다. 연구에 사용된 마커 유전자는 기존 연구에 밝혀진 마커 유전자 중 gbbct 파일에 나타나는 16S rRNA, hsp65, rpoB, sodA, tuf를 이용하였다(Springer et al., 1995; McNabb et al., 2004; Devulder et al., 2005; Mignard et al., 2007; Adekambi et al., 2006).

NTM 44종의 모든 유전 정보와 accession을 확인하기 위한 검색 기능을 위해 accession-organism-definition-cds-gene-product 6항목의 형태로 나열되는 데이터베이스를 구축하였다. 이는 NTM 44종 중 연

구자가 원하는 5개 마커 유전자를 포함한 다른 모든 유전자 명칭을 입력하여 해당하는 accession과 구체적인 설명이 있는 definition 항목을 검색하기 위함이다. 테이블 형태는 모두 문자형으로 varchar 타입으로 저장하였다. 검색 결과로 확인된 definition과 accession을 이용하여 해당 유전자의 sequence를 검색하고 분석에 사용 가능하도록 fasta 형식의 데이터를 종별 또는 마커 유전자 별로 다운 가능하도록 하였다. 기호 “>”로 시작하는 fasta 형식의 데이터는 organism-accession-gene-product-sequence의 항목을 나타내고 44종의 데이터와 5개의 마커 유전자 별로 연구자가 내려 받기 가능하도록 구현하였다. 내려 받은 파일들은 웹 서버에서 제공하는 분석 프로그램인 ClustalX, MEGA, SequenceMatrix를 이용하여 사용자가 분석 가능하다. table 2.1은 데이터베이스에 사용된 5가지 마커 유전자와 MySQL 데이터베이스 테이블 형식이다.

본 데이터베이스는 NTM 종과 그에 대한 모든 유전자 정보, 마커 유전자 서열 검색 등 분석에 활용될 데이터를 지원하기 위해 구축되었으므로 연구자가 유전자 서열을 찾고자 할 경우, 우선 accession, organism, gene, product 명칭을 입력하여 원하는 유전자에 대해 특화된 정보를 검색할 수 있다. 각 항목을 원하는 만큼 검색에 이용할 수 있고, 또한 임의의 종이 갖고 있는 특정한 gene을 찾을 경우는 원하는 종과 gene 정보를 입력하여 교차 검색을 통해 가능하도록 하였다. 검색어를 정확히 알고 있지 못할 경우에도, 해당 항목에 입력 단어를 포함하고 있는 데이터를 모두 보여줄 수 있도록 하였다. 이를 통해 검색된 종의 유전자에 대한 자세한 설명과 accession 번호도 출력되어 서열 정보 검색에 사용할 수 있도록 하였다. 그리고 본 데이터베이스의 NTM 44종과 5개의 마커 유전자 서열

정보는 fasta 형식으로 종별, 마커 유전자 별로 다운로드 가능하도록 열거하였다. 연구자는 원하는 종의 모든 유전자 정보와 서열 정보, 원하는 마커 유전자를 지닌 종의 유전자 서열 정보를 선택해서 fasta 형식으로 다운로드 가능하다.

연구자가 본 데이터베이스로부터 서열 정보 데이터를 받고 싶을 경우는 검색한 유전자의 accession을 입력하여 해당 유전자의 sequence를 fasta 형식으로 연구자가 다운 받을 수 있게 하였다. 서열 정보 검색은 accession을 fasta 파일 다운로드 페이지에 입력하면 해당 accession의 유전 서열 정보를 fasta 형식으로 다운 받을 수 있다. 또한 organism, gene, product 입력으로 fasta 형식의 유전 서열 정보 다운로드가 가능하며, 종별 유전자 서열 데이터, 유전자 별 서열 데이터의 수집이 가능하다. ClustalX (Larkin et al., 2007), MEGA (Kumar et al., 1994), SequenceMatrix (Vaidya et al., 2011) 서열 정렬 및 분석 프로그램을 통해 마커 유전자의 조합을 통한 계통수를 작성하여 NTM 동정에 사용할 수 있도록 하였다. 또한 NTM 종의 유전 정보에 대해 다양한 옵션으로 검색할 수 있고 fasta 파일 형식으로 서열 정보를 다운 받을 수 있다. 다운로드한 파일들은 웹 서버에서 제공하는 setup 파일들을 통해 설치되는 분석 프로그램을 이용한 연구에 활용 가능하도록 하였다. 본 데이터베이스는 다양한 마커 유전자 조합을 통한 동정법 연구의 의의를 보이고 연구자가 각자의 연구에 필요한 프로그램과 데이터를 구축할 수 있도록 지원하고자 한다.

2.3 서열 정렬 및 분석 프로그램

근래의 박테리아 분석에는 복수의 마커 유전자 조합을 통한 방법이 많이 이용되고 있다. 특히, 16S rRNA를 포함한 4~5가지 마커 유전자를 이용한 동정법이 권장되고 있다. 하지만 두 유전자가 분류 계통도에서 서로 다른 단계에서 유용하다면 두 유전자의 조합은 결국 효과적이지 못하다. 이러한 점에서 NTM의 경우는 모두 같은 속 (genus)에 속하기 때문에 조합하는 마커 유전자가 개수가 증가할수록 정확한 계통 분류가 가능할 것으로 기대된다. 본 연구에서는 문헌 조사를 통해 알려진 마커 유전자들 중 NCBI gbbct 파일 데이터에 20종 이상의 NTM종이 보유하고 있는 마커 유전자 5개를 대상으로 하였다. 그 중 마커 유전자 4개와 5개 조합을 이용한 동정법을 비교하기로 한다. 이를 통해 다양한 유전자 조합이 동정법에 어떠한 영향을 미치는지 알아보고자 한다.

본 데이터베이스는 연구자가 원하는 NTM 종과 유전자에 대한 정보를 제공하여 유전자 재조합을 통해 NTM의 생명정보학적 동정이 가능하도록 하고자 한다. 이를 위해 본 데이터베이스에서 다운로드한 fasta 파일을 이용하여 서열 정렬 및 계통 분류도 작성이 가능하도록 최근 많이 쓰이는 분석 프로그램을 함께 제공한다. ClustalX는 서열 정렬 프로그램으로 서열 간에 비교 분석이 가능하도록 길이와 공통된 서열 구간의 위치가 다른 비교 대상의 서열들을 정렬해 주는 프로그램이다. Fasta 형식의 파일 입력이 가능하여 본 데이터베이스로부터 다운로드한 데이터를 입력하여 서열 정렬이 가능하다. ClustalX를 통해 서열 정렬된 NTM 종의 마커 유전자 별 서열을 유전자 조합 프로그램을 이용해 조합할 수 있다.

SequeunceMatrix는 서열 조합 프로그램으로 종별로 복수의 마커 유전자를 조합하여 비교할 수 있도록 하는 프로그램이다. 정확한 비교 분석을 위해서는 마커 유전자 별로 종의 이름과 개수가 일치하여야 한다. SequenceMatrix를 통해 조합된 마커 유전자는 파일로 저장하여 MEGA 프로그램을 이용하여 분석한다. MEGA를 통해 각 분석법의 종별 pair-wise 분석, 평균 pair-wise 분석 수치, 계통 분류도 작성이 가능하다. 이를 통해 각 유전자 조합별 분석법의 동정 변별력과 견고함을 분석할 수 있다. Table 2.2는 분석 용도에 따른 프로그램 목록이다.

```

LOCUS      CP001682          1617804 bp    DNA    circular BCT 28-JAN-2014
DEFINITION Cryptobacterium curtum DSM 15641, complete genome.
ACCESSION  CP001682 ABTM01000000 ABTM01000001 ABTM01000002
ABTM01000003
VERSION    CP001682.1 GI:256582603
DBLINK     BioProject: PRJNA20739
           BioSample: SAMN02598434
KEYWORDS
SOURCE     Cryptobacterium curtum DSM 15641
ORGANISM   Cryptobacterium curtum DSM 15641
           Bacteria; Actinobacteria; Coriobacteridae; Coriobacteriales;
           Coriobacterineae; Coriobacteriaceae; Cryptobacterium.
REFERENCE  1 (bases 1 to 1617804)
AUTHORS    Mavrommatis,K., Pukall,R., Rohde,C., Chen,F., Sims,D., Brettin,T.
TITLE      Complete genome sequence of Cryptobacterium curtum type strain
           (12-3)
JOURNAL    Stand Genomic Sci 1 (2), 93-100 (2009)
PUBMED     21304644
REMARK     Publication Status: Online-Only
FEATURES   Location/Qualifiers
           source          1..1617804
                           /organism="Cryptobacterium curtum DSM 15641"
                           /mol_type="genomic DNA"
                           /strain="DSM 15641"
                           /culture_collection="DSM:15641"
                           /db_xref="taxon:469378"
                           /note="type strain of Cryptobacterium curtum DSM 15641"
           gene            2977..4113
                           /locus_tag="Ccur_00030"
           CDS              2977..4113
                           /locus_tag="Ccur_00030"
                           /note="PFAM: RecF/RecN/SMC N terminal domain;
                           TIGRFAM: recF protein"
                           /product="recF protein"
                           /protein_id="ACU93740.1"

/translation="MPSLEEPDKNLRLTNLVLRNFRNHQEFSLKGLQGITILAGPNAT
GKTSVVEAIQLISALKSFRASQIGRAIRWGQTAASVIATIESDHRQLDLQLRIIEGKR"
ORIGIN
1   ttagtgtata cgttattct tttacctac tgctgaagac tatgaaaagc ttataggtaa
61   aggcattccac gtttagccca acttctgat acactacacc agtgcacgga acatggtgaa
121  tgagggttta tgtcagcaga taaatatgac gcgctgtgga atacagttgt tgaacgcatt
181  aaaacatatg aaaaattga tgcattccag atagatgcgt tctttctcg acttggtcct

```

Figure 2.1 “gbtseq” file format. “gbtseq” files in GenBank provide the detailed information on bacteria.

a) Database format

```

|AB239921      Mycobacterium arupense      Mycobacterium arupense rpoB gene for
RNA polymerase beta subunit, partial cds, strain:CST7052 (= GTC2730),      <1..>
398      rpoB      RNA polymerase beta subunit      BAF02281.1
TVGELIQNQRVGLSRMERVVRERMTTQDVEAITPQT LINIRPWAAIKEFFGTSQLSQFMD
QNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHP SHYGRMCPIETPEGPNIGLIGSLAT
YARVNPFFGF
accgtggcgagctgatccagaaccagatccgggtcgggtctccggatggagcgcgtggtccgcgagcg
gatgaccaccaccaggagctgcgaggccatcacgcccagaccctgatcaacatccgcccggtggtcgcgcg
atcaaggagttcttcggcaccagccagctctccagttcatggaccagaacaacccgctgctgggtcgtgacc
acaagcgccgctgctgcggcgtgggcccggcgtgctgcgctgagcggccgggctggaagttcgtgac
gtgcaccgctccactacggccggatgtgctcgatcgagacccgggaaggcccaacatcggttgatcggg
tcgctgacgacatgacgcccgggtcaacccgttcgggtcat
AB239922      Mycobacterium arupense      Mycobacterium arupense hsp65 gene
for 65 kDa heat shock protein, partial cds, strain:CST7052 (= GTC2730),      <1..>
424      hsp65      65 kDa heat shock protein      BAF02282.1
GVSLAKEIELEDPYKIGAEKVKEVAKTDDVAGDGT TTTATVLAQALVKEGLRNVAAGANPL
GLKRIGIEKAVAKTEGLLATAKAIE TKDQIAATAGISAGDQTIGDLIAEAMDKVGNEGVTVEE
SNTFGLQLELTEGMR
atggtgtgtccatcgccaaggagatcgagctggaggaccgctacgagaagatcgggcgccgagctggtcaaa
gaggtcgccaaagaagaccgagcagctgcgggtgacggcaccaccaccgcccagctgctgcggcggcc
ctggtcaaggaagcgtcgcaacgtggccgcccggcgccaacccgctgggtctgaagcggcgtcgcgaga
aggcgtgcgcaagatcaccgaaggtctgctgacacggccaaggcagatcgagaccaggaccagatcg
ggctacggccggtatctccgcccggcgaccagaccatcggtgacctgacgcccaggccatggaacaaggtcg
gcaacgaggggtgctacacgctcgaggagtcacacaccttcggcctcgagctggagctcaccgagggatgc

```

b) Fasta format

```

>Mycobacterium_abscessus|AB548598 | hsp65 | heat shock protein 65

CGCCAAGGAGATCGAGCTGGAGGACCCGTACGAGAAGATCGGCGCTGAGCTGGTCAAG
GAAGTTGCCAAGAAGACCGACGACGTCGCGGGTGACGGCACCCACCACCGCCACCGTTC
TTGCCAAGGCCCTGGTCAAGGAAGGCTGCGTAACGTGCGCGCCGCCGCAACCCGCT
CGGCCTGAAGCGCGGCATCGAGAAGGCCGTGAGAAGGTCACCGAGACGCTGCTGAAG
AGCGCCAAGGAGGTCGAGACCAAGGAGCAGATCGCGCCACGGCCGGTATCTCCGCG
GGCGACCAAGTCCATCGGCCACCTGATCGCCGAGGCCATGGACAAGGTTGGCAACGAGG
GTGTCATCACCGTCCGAGGAGTCCAACACCTTCGGCCTGCAGCTGGAGCTCACCG
>Mycobacterium_arupense|JN571190 | hsp65 | 65 kDa heat shock protein

GCCAAGGAGATCGAGCTGGAGGACCCGTACGAGAAGATCGGCGCCGAGCTGGTCAAG
AGGTCGCCAAGAAGACCGACGACGTCGCGGGTGACGGCACCCACCACCGCCACCGTGT
GGCCACAGGCCCTGGTCAAGGAAGGCTGCGCAACGTAGCCGCCGCGGCCAACCCTGCT
GGGTCGCAAGCGCGGCATCGAGAAGGCCGTGCAAGATCACGGAAGGCTGCTTGC
ACGGCCAAGGCGATCGAGACCAAGGAGCAGATCGCGGCTACGGCCGGTATCTCCGCG
GCGACCAAGACATCGGTGACCTGATCGCCGAGGCCATGGACAAGGTCGCGCAACGAGGG
TGTCATCACCGTCCGAGGAGTCCAACACCTTCGGCCTGCAGCTGGAGCTCACCG
>Mycobacterium_arupense|JQ617911 | hsp65 | 65 kDa heat shock protein

CCCGTACGAGAAGATCGGCGCCGAGCTGGTCAAGAGGTCGCCAAGAAGACCGACGAC
GTCGCGGGTGACGGCACCCACCACCGCCACCGTGTCTGCGCCAGGCCCTGGTCAAGGAA
GGCTGCGCAACGTGGCCGCCGCGGCCAACCCTGCGGTCTGAAGCGCGGCATCGAG
AAGGCCGTGCGCAAGATCACCGAAGGCTGCTTGCCACGGCCAAGGCGATCGAGACCA
GGACCAAGATCGCGGCTACGGCCGGTATCTCCGCGGGCGACCAAGACATCGGTGACCTG
ATCGCCGAGGCCATGGACAAGGTCGCGCAACGAGGGTGTCATCACCGTCGAGGAGTCCA
ACACCTTCGGCCTGCAGCTGGAGCTCACCGAGGGCATGCGCTTCGACAAAGGCTACATC
TCGGGCTACTTCGTCACCGACGCCGAGCGTCAAGGAGCCGCTCTGGAGGACCCATCAT
CCTGCTGGTCAGCTCGAAGATCTGACCGTCAAGGACCTGCTGCCCTTGTGGAGAAGG
TCATCCAGTCCGGCAAGCCGTTGCTGATCATCGCCGAGGACGTCGAGGGCGAAGCCCT
GTCGACCCCTGGTGGT

```

Figure 2.2 Database format and Fasta format. Java language program has parsed the gbbct.seq files into the data; a) database format b) fasta format.

Table 2.1 Marker gene list and database table structure

a) Marker gene list

Name of marker gene	Description
16S rRNA	Bacterial ribosomal RNA
Hsp65	Heat shock protein 65
rpoB	RNA polymerase β -subunit
sodA	Manganese Superoxide dismutase
Tuf	Elongation factor tu

b) Database table structure

Name of column	Type of data
ACCESSION	Varchar(255)
ORGANISM	Varchar(255)
DEFINITION	Varchar(800)
CDS	Varchar(255)
GENE	Varchar(255)
PRODUCT	Varchar(255)

a) 5 marker genes are used in the analysis. b) The database for searching is constructed by 6 column

Table 2.2 Analysis programs

Analysis	Program
Sequence alignment	ClustalX 1.83
Sequence combination	SequenceMatrix 1.7.8
Pair-wise analysis & phylogenetic analysis	MEGA 6.0
blast	WWWBlast2.2.30+

Four programs are needed for sequence alignment, combination, analysis.

제 3장. 연구 결과

3.1 NTM 종과 마커 유전자에 대한 검색 시스템 구축

본 연구에서는 44종의 NTM에 대한 유전 정보와 5개의 마커 유전자의 서열 정보를 검색할 수 있도록 데이터베이스를 구축하였다. 각 종에 대한 유전 정보는 마커 유전자를 포함한 GenBank에 존재하는 모든 유전자에 대한 정보를 담고 있고 “NTM Gene Search” 페이지에서 검색을 통해 accession을 확인하여 “Marker-sequence Searching” 페이지에서 서열 정보를 검색하여 다운 받거나 직접 organism, gene, 또는 product 명을 검색하여 다운로드 할 수 있다. “Sequence Fasta Files” 페이지에서는 44종이 각각 보유하고 있는 모든 유전 정보와 5개의 마커 유전자 서열 정보를 직접 다운로드 할 수 있도록 설정하였다. “Blast & Analysis Programs” 페이지에서는 ClustalX, SequenceMatrix, MEGA 프로그램을 링크하여 연구자가 다운로드한 데이터를 이용하여 직접 서열 정렬, 조합, 분석할 수 있도록 하였다. 또한 NCBI의 BLAST 웹 링크를 통하여 연구자가 상동성 검색이 필요할 경우 사용할 수 있도록 하였다. 44종에 대한 유전 정보는 종마다 천차만별이었다. 가장 적은 정보를 보유한 *M. shinjukuense*는 6개의 유전자 서열 정보만을 나타낸 반면, *M. bovis*는 가장 많은 28461개의 유전자 서열 정보를 보유하는 것으로 나타났다. 5개의 마커 유전자 중에서는 16S rRNA가 가장 많은 유전자 서열 정보인 1352개를

가지고 있었으며, 이 밖에 Hsp65는 828개, sodA는 55개, tuf는 69개, rpoB는 841개 유전자 서열 정보를 나타냈다. 서열 정보가 많을수록 마커 유전자로서 활용도가 높다. Table 3.1 3.2는 유전자 별 데이터와 종 별 데이터 보유 현황을 나타낸 것이다.

앞에서 간단히 언급하였듯이 본 데이터베이스는 총 5개의 웹 페이지로 구성되어 있다. Figure 3.1 “Home” 페이지에는 NTM과 본 시스템에 대한 간단한 설명과 함께 각 페이지를 나타낸다. “NTM Gene Search” 페이지에서는 accession, organism, gene, product 중 원하는 항목에 선택하고 입력하면 그에 해당하는 유전 정보가 6항목으로 출력된다. 두 개 이상의 항목을 입력할 경우는 각 항목의 입력된 정보가 공통으로 일치하는 정보를 출력한다. Figure 3.2에서 organism과 gene 입력란에 각각 “Mycobacterium aubagnense”와 “gyrB”를 입력하고 있고 Figure 3.3는 검색 결과이다. 이를 통해 M. aubagnense의 gyrB 유전자의 accession 넘버는 AM421322이고 product는 DNA gyrase subunit B라는 결과가 나왔다. definition 항목에서 해당 유전자는 Mycobacterium aubagnense 중 CCUG50186T 균주의 DNA gyrase subunit B 단백질을 위한 gyrB 유전자라는 것을 알 수 있다. CDS는 전체 유전자에서의 해당 유전자 서열이 위치하는 서열 좌표이다. 검색 결과로 본인이 원하는 유전자의 찾아 해당 유전 정보의 세부 정보와 accession 넘버를 알 수 있다. “NTM Gene Search” 페이지에서 검색된 accession 넘버 AM421322를 “Marker-sequence Searching” 페이지에 입력하게 되면 figure 3.4처럼 M. aubagnense의 유전자 gyrB accession 넘버 AM421322에 대한 유전자 서열 검색 결과를 다운로드 할 수 있다. 이처럼 사용자는 “NTM Gene Search” 페이지에서 유전자에 대한 대략적인 검색을 통해 accession, definition, CDS 좌표와

같은 상세한 정보를 확인하고 이를 “Marker-sequence Searching” 페이지에 입력함으로써 Figure 4와 같이 유전자 서열 정보를 다운로드 할 수 있다. 마커 유전자뿐만 아니라, 44종을 대상으로 한 본 데이터베이스의 범위 안에서 원하는 유전자에 대한 모든 유전자 서열 정보를 검색을 통해 다운로드 할 수 있다. Figure 3.5는 “dnaJ”라는 검색어를 통해 유전자 dnaJ를 가지고 있는 종에 대한 유전자 서열 정보를 다운로드한 결과이다. “Sequence Fasta Files” 페이지에서는 기본적으로 본 데이터베이스에 포함되어 있는 NTM 44종 각각의 모든 유전자 서열 정보와 마커 유전자 5개의 유전자를 기준으로 한 유전자 서열 정보를 파일로 로드 할 수 있도록 하였다. 이를 이용하여 연구자는 원하는 종의 모든 유전자 서열 정보와 마커 유전자에 대한 서열 정보를 얻을 수 있다. 다운로드한 파일은 연구자가 임의로 편집하여 원하는 종과 서열만을 추출하여 분석에 활용할 수 있다. Figure 3.6는 sodA 유전자 파일을 다운로드한 결과이다. Figure 3.7의 “Blast & Analysis Programs” 페이지에서는 연구자가 다운로드한 서열 정보 파일을 서열 정렬, 유전자 조합, pair-wise 분석, 계통 분류도 분석할 수 있도록 서열 정렬 분석 프로그램을 다운로드 할 수 있도록 링크하였다. ClustalX는 분석이 가능한 형태로 대상 서열들을 정렬하는 프로그램이다. SequenceMatrix는 정렬된 서열들을 유전자 별로 조합하여 분석 가능한 형태로 출력하는 프로그램이다. MEGA는 정렬 후 조합된 서열들을 pair-wise 분석과 계통 분류도 분석하는 프로그램이다. 연구자들은 이를 이용하여 정렬된 유전자 서열들의 다양한 조합을 통해 다양한 분석을 할 수 있다. NCBI의 BLAST 웹 사이트를 링크하여 유전자 서열의 상동성 분석도 가능하다.

3.2 NTMDB에 의한 마커 유전자 조합 계통분류 분석

본 연구에서는 NTMDB를 이용하여 추출한 16S rRNA, hsp65, rpoB, sodA, tuf 각각의 마커 유전자를 이용한 계통 분류도와 기존 연구와의 비교를 통해 각 마커 유전자의 활용도를 확인하였다. 그 후에 기존에 연구되었던 16S rRNA, hsp65, rpoB, sodA를 조합한 계통 분류도와 16S rRNA, hsp65, rpoB, sodA에 tuf까지 포함하여 조합한 계통 분류도를 작성하고 비교 분석하였다. 이를 통해 NTM 연구에 다양한 마커 유전자 조합이 이용될 필요성에 대해 제시하고자 하였다. 각 유전자 서열 정렬은 clustalX를 이용하였고, MEGA를 통해 계통 분류도 및 분석을 시행하였다. 비교할 기존 연구(Rogall et al., 1990; Devulder et al., 2005) 방법과 같이 Neighbor-joining 방식의 분류 계통도의 Kimura 2-parameter 모델로, bootstrap replication은 500으로 설정하였다. Neighbor-joining은 가장 격리된 근접군 쌍 사이와 나머지 가지를 연속해서 삽입하고 가장 가까운 근접군 쌍을 통합하는 방식으로 진화적 순서를 중점으로 한 Maximum-likelihood 분석보다는 계산이 빠르고 염기서열간의 차이를 거리로 환산한 대표적인 거리기반 방식이다(Saitou et al., 1987). kimura-2-parameter model은 4가지의 염기가 같은 주기로 나타나고 위치에 상관없이 같은 비율로 치환된다는 가정 하에 transition과 transversion 두 가지 종류의 염기 치환을 구분하는 모델이다. Transition은 아데닌(adenine)과 구아닌(guanine)으로 구성된 퓨린(purine)간 또는 시토신(cytosine)과 티민(thymine)으로 구성된 피리미딘(pyrimidine)간의 4가지 경우의 치환을 말한다. transversion은

퓨린과 피리미딘간의 8가지 경우의 치환을 의미한다(Kimura, 1980).

16S rRNA 서열을 이용한 계통분류의 정확성을 알아보기 위해 본 44종의 NTM 데이터베이스 데이터 중 기존 연구(Devulder et al., 2005)와 공통의 종인 33종을 추려서 계통 분류하여 figure 3.8을 작성 하였다. 기존 연구와 figure 3.8을 비교한 결과, 전반적으로 NTM이 발육 속도에 따라 위쪽은 완속발육군, 아래쪽은 신속발육군으로 분류가 이루어진 것을 볼 수 있다. *M. gilvum*, *M. parafortuitum*; *M. ulcerans*, *M. marinum*; *M. avium*, *M. paratuberculosis*; *M. chelonae*, *M. abscessus*의 유전적 연관성을 나타낸 것도 확인할 수 있다. 하지만, *M. lentiflavum*, *M. interjectum*, *M. gordonae*, *M. celatum*, *M. triviale*은 기존 연구에서 완속발육군이었지만, figure 3.8에서는 신속발육군으로 분류되었다. 반면에 *M. elephantis*는 신속발육군이지만 완속발육군으로 분류되었다. 마이코 박테리아 동정법에 대한 기존 연구(Rogall et al., 1990; Devulder et al., 2005)에서는 bootstrap 50%와 80% 이상인 node를 비교하였지만 본 연구에서는 대상 NTM종이 한정적인 관계로 bootstrap 50%와 75%를 비교하였다. 그 결과, 30개의 node 중 bootstrap 치수가 75% 이상인 node는 8개, 50% 이상인 node는 18개였다. hsp65의 계통분류도 16S rRNA 방법과 같은 방식으로, NTM 44종 중 기존 연구(Devulder et al., 2005)의 mycobacteria 종들과 공통인 33종을 추려서 작성한 figure 3.9를 비교하였다. hsp65에 의한 계통 분류를 한 결과 *M. avium*, *M. paratuberculosis*, *M. intracellulare*; *M. marinum*, *M. ulcerans*; *M. gilvum*, *M. parafortuitum*의 분류적 연관성이 잘 나타나 있다. 그 외에 *M. celatum*, *M. kansasii*, *M. shimoidei* 군집도 기존 연구와 유사하게 형성되었고 발육 속도에 의한 구분도 잘 되어 있다. 30개의 node 중 75% 이상의 bootstrap을 나타낸 node는 6개, 50% 이상

인 node는 11개였다. *rpoB* 유전자 역시 기존 연구(Devulder et al., 2005)와 공통인 NTM 33종으로 작성한 figure 3.10을 기존 연구와 비교하였다. *M. marinum*, *M. ulcerans*; *M. paratuberculosis*, *M. avium*; *M. goodii*, *M. smegmatis*의 3개 군집의 연관성이 일치하였고, 발육속도에 의한 구분도 일치하였다. 기존 연구와 비교하면 *M. chelonae*, *M. abscessus*; *M. terrae*, *M. nonchromogenicum*; *M. neoaurum*, *M. peregrinum*의 연관성을 확인할 수 있다. 30개의 node 중 75% 이상의 bootstrap을 나타낸 node는 5개, 50% 이상인 node는 9이었다. *sodA* 계통 분류는 기존 연구(Devulder et al., 2005)와 figure 3.11의 공통된 NTM 19종을 대상으로 비교하였다. 발육속도에 의한 구분이 잘 되어 있었다. *M. marinum*, *M. ulcerans*; *M. avium*, *M. paratuberculosis*; *M. wolinskyi*, *M. goodii*, *M. smegmatis*의 연관성을 볼 수 있다. 16개의 node 중 8개가 bootstrap의 치수가 75%를 넘었고, 12개 50%를 넘었다. *tuf* 유전자 계통 분류를 위해 기존 연구(Rogall et al., 1990)와 figure 3.12의 36종의 NTM을 비교하였다. *M. phlei*, *M. elephantis*, *M. parascrofulaceum*을 제외하고는 발육 속도에 의한 구분은 잘 이루어졌다. *M. avium*, *M. paratuberculosis*, *M. intracellulare*; *M. abscessus*, *M. bolletii*, *M. chelonae*; *M. gilvum*, *M. parafortuitum*; *M. neoaurum*, *M. phocaicum*, *M. aubagnense*, *M. mucogenicum*의 분류적 연관성을 확인할 수 있다. *M. smegmatis*, *M. goodii*, *M. wolinskyi*; *M. terrae*, *M. nonchromogenicum*, *M. triviale*; *M. marinum*, *M. ulcerans*는 어느 정도 연관성을 보이며 군집을 형성하였다. 75% 이상의 bootstrap을 나타낸 node는 32개중 7개였고, 50% 이상의 node는 17개였다. 각각의 단일 마커에 대한 분석 결과는 16S rRNA의 계통 분류의 일부를 제외하면 16S rRNA, *hsp65*, *rpoB*, *sodA*, *tuf* 각각에 의한 계통 분류는 발육 속도에 의한 분류가 잘 이루어

졌다. 하지만 각각의 계통 분류에서 bootstrap 수치가 80% 이상인 node가 21.01% 이었다. 또한 16S rRNA 유전자에 의한 계통 분류도에서 *M. partuberculosis*, *M. avium*와 *tuf* 유전자에 의한 계통 분류도에서 *M. abscessus*, *M. bolletii*와 같이 분류적 연관성은 있으나 구분하지 못하는 경우도 있는 것으로 나타났다. 이로써 다섯 유전자가 지금까지 여러 연구에서 마커 유전자로 제시되어 왔지만, 어느 하나만의 유전자로는 정확도에 한계가 있으며, 일관성 있는 계통 분류가 어렵다는 것을 알 수 있다.

이제 본 데이터베이스의 데이터를 이용하여 16S rRNA, *hsp65*, *rpoB*, *sodA*를 조합한 계통 분류도와 16S rRNA, *hsp65*, *rpoB*, *sodA*, *tuf*를 조합한 계통 분류도를 비교하여 조합에 따라 분류 계통도의 분별력과 견고성을 확인하고자 한다. 분석 대상에 대한 유전자를 clustalX를 이용하여 서열 정렬 후 SequenceMatrix를 이용하여 유전자를 조합하였고 MEGA를 통해 계통 분류도를 작성하였다. figure 3.13와 3.14는 각각 참고 분류도와 비교한 16S rRNA, *hsp65*, *rpoB*, *sodA*, *tuf*의 종들 중 공통으로 보유하고 있는 16종을 대상으로 16S rRNA, *hsp65*, *rpoB*, *sodA*를 조합한 계통 분류도와 16S rRNA, *hsp65*, *rpoB*, *sodA*, *tuf*를 조합한 계통 분류도이다. 두 분류도를 비교한 결과 발육 속도, 종들 간의 연관성이 거의 일치하였다. *M. marinum*, *M. ulcerans*; *M. avium*, *M. paratuberculosis*; *M. abscessus*, *M. chelonae*; *M. wolinskyi*, *M. goodii*, *M. smegmatis*의 분류적 연관성이 기존 연구와 일치하였다. 그 외에 *M. kansasii*, *M. bovis*; *M. mucogenicum*, *M. fortuitum*이 기존연구와 같이 군집을 형성한 것을 볼 수 있다. 반면에 Figure 3.13는 *M. mucogenicum*이 먼저 진화상으로 분리되고, *M. fortuitum*과 *M. wolinskyi*-*M. goodii*-*M. smegmatis* 그룹 순으로 분리된 것으로 분석되

었다. 하지만 Figure 3.14는 *M. mucogenicum*과 *M. fortuitum*이 *M. wolinskyi*-*M. goodii*-*M. smegmatis* 그룹으로부터 먼저 분리되고 그 뒤로 각각 진화된 것을 확인 할 수 있다. *tuf* gene 유전자를 포함함에 따라 다른 진화적 계통 분류도가 생성된 것이다. 두 조합은 bootstrap 치수에서도 약간의 차이를 보였다. 16S rRNA-hsp65-rpoB-sodA 조합의 경우 50% 이상의 bootstrap 치수를 나타내는 node가 13개 중 11개를 나타낸 것에 비해, 16S rRNA-hsp65-rpoB-sodA-tuf 조합에서는 13개의 node 중 13개 모두 50% 이상의 bootstrap 치수를 보였다. 반면에 75 이상의 bootstrap 치수는 두 조합 모두 13개 node 중 10개로 나타났다. 결과적으로 50 이상의 bootstrap 치수를 가지는 node를 다소 더 많이 보유한 16S rRNA-hsp65-rpoB-sodA-tuf의 조합이 조금 더 견고한 계통 분류도를 작성할 수 있음을 의미한다. 두 조합의 분류도의 특징 중 하나는 bootstrap 치수 100을 나타내는 node가 4개가 있어서 bootstrap 치수 100의 node가 없거나 적은 다른 단일 유전자에 의한 분류보다 견고함을 확인 할 수 있다는 것이다. 두 조합의 최소 bootstrap가 각각 40, 58인 반면에 단일 유전자에 의한 분류는 1부터 다양한 것으로 나타났다.

Figure 3.15는 16S rRNA-hsp65-rpoB-sodA 조합과 16S rRNA-hsp65-rpoB-sodA-tuf 조합으로 각 종간에 p-distance pairwise matrix 결과이다. p-distance는 두 비교 서열 간에 다른 DNA 서열이 전체 서열 중에 어느 정도 차지하느냐를 수치로 표현한 것으로 계통 분류도의 변별력을 나타낸다. p-distance 수치가 낮으면 두 서열 간에 차이가 없고 수치가 높으면 차이가 많이 나는 것이다. 두 matrix는 수치가 조금씩 차이 날 뿐 큰 차이가 없었다. 이처럼 두 조합에 의한 분류 계통도 형태와 pairwise matrix의 수치 분포가 상당히 유사하였고 이

는 두 조합에 의한 종 분류가 상당히 일관성 있음을 알 수 있다. Table3.3은 각 유전자와 조합 별 분석 수치를 나타낸 것이다. 이에 의하면 유전자 별 G+C 비율은 57~67%로 mycobacteria 속(genus)의 평균 범주와 일치한다. 평균 p-distance와 평균 kimura-2-parameter model distance를 보면 각 유전자의 수치는 제각각 이지만, 16S rRNA-hsp65-rpoB-sodA 조합과 16S rRNA-hsp65-rpoB-sodA-tuf 조합의 비교에서는 5개 조합인 경우에 각각의 값이 0.001씩 차이를 나타내어 16S rRNA-hsp65-rpoB-sodA-tuf 조합이 계통 분류적인 분별력이 상승된 것을 알 수 있다. 두 조합의 bootstrap 수치에서는 75% 이상의 비율은 같지만 50% 이상의 비율은 16S rRNA-hsp65-rpoB-sodA-tuf 조합이 100% 이기 때문에 계통 분류도의 견고함이 상승된 것을 볼 수 있다. sodA의 경우 p-distance와 K2P 수치가 두 조합에 비해 높았으나 이는 average fragment size가 비교적 작고 50%와 75% 이상의 bootstrap 수치도 낮기 때문에 두 조합과 비교하기에는 견고함이 낮다고 볼 수 있다. 두 조합의 p-distance와 K2P distance, bootstrap을 비교하면 16S rRNA-hsp65-rpoB-sodA-tuf 조합이 16S rRNA-hsp65-rpoB-sodA 조합에 비해 변별력과 견고함에 있어서 더 안정된 계통 분류도 모델임을 알 수 있다. 이를 통해 정확한 종 동정을 위해서는 더 많은 마커 유전자의 발견이 필수적이고 이들을 조합하여 분석하는 동정 연구가 이루어져야 함을 확인할 수 있다.

3.3 WWWBlast 분석 결과

WWWBlast는 연구자 고유의 서버에 데이터베이스를 갖추고 NCBI의 blast 분석을 실행할 수 있는 프로그램이다. 본 연구에서는 NTM 44종의 16S rRNA, hsp65, rpoB, sodA, tuf 염기 서열에 대한 데이터베이스를 구축하여 *M. peregrinum*을 기준으로 blastn과 tblastx 분석으로 상동성 검색을 실시하였다. Figure 3.16은 PC command prompt창에서 NTM 44종의 16S rRNA 유전자에 대한 database를 구축하고 *M. peregrinum* 종의 16SrRNA 염기 서열을 query로 blastn과 tblastx를 실행하는 과정을 나타낸다. Figure 3.17은 *M. peregrinum* 종의 blastn 결과 파일 중 query가 gene ID = AF058712인 부분이다. bit score와 e-value 순으로 NTM 종과 유전자 정보가 나열되고, 그 밑에 query와 subject 서열의 identity를 나타낸다. Blastn은 염기 서열간의 비교를 위한 분석법인 반면, tblastx는 입력한 서열과 염기서열 database를 모두 6개의 frame으로 변환 후 비교하는 분석법이다. “bit score” S' 는 alignment 각각의 letter-to-letter 와 letter-to-null score를 더하여 계산된 raw score, S 를 통계 변수인 K 와 λ 를 이용하여 정규화한 scoring system이다. 따라서 bit score는 size와 search space에 관계없이 서로 다른 scoring system을 사용한 alignment 의 score간에 비교가 가능하다. score가 높을수록 유사도가 높다. E-value는 주어진 서열의 정렬이 유의할 가능성을 나타내는 통계적 수치이다. 정렬이 우연히 발생할 수 있는 확률로 0에 가까울수록 유의하다.

$$S' = \frac{\lambda S - \ln K}{\ln 2}$$

S: raw score

K and lambda: statistical parameters of the scoring system depending on substitution matrix and on the gap penalties

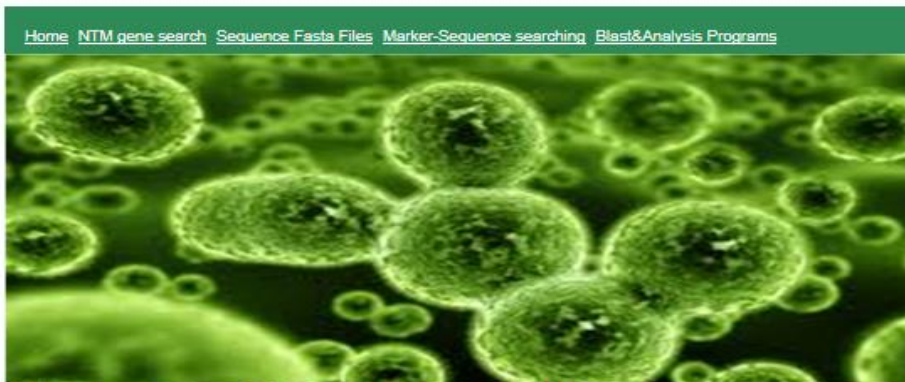
Table 3.4는 WWWBlast를 이용하여 NTM 44종의 5개 마커 유전자 염기 서열에 대한 데이터베이스를 구축하여 실시한 blast 분석에 수치를 나타낸다. gap은 두 염기 서열 정렬 시 null과 상동하는 염기서열 위치를 의미한다. Table 3.4에서는 gap를 고려하는지에 대한 여부에 따라 blastn의 lambda 값을 나타내고 있다.

Blastn과 tblastx의 결과 파일은 WWWBlast에 구축된 데이터베이스의 모든 염기 서열에 대한 분석 결과이기 때문에 JAVA language를 이용하여 각 마커 및 blast 결과에 따른 종별 리스트 파일을 작성하였다. Figure 3.18, 3.19는 특정 *M. peregrinum*의 16S rRNA 염기 서열을 query로 하여 blastn과 tblastx를 실행한 결과를 bitscore가 높은 종별로 추출한 것이다. 16S rRNA의 경우 MEGA figure에서도 확인한 바와 같이 blastn과 tblastx에서 모두 *M. wolinskyi*의 염기 서열이 *M. peregrinum*의 염기서열과 가장 유사한 것으로 나타났다. 그 외에 *M. gilvum*, *M. neoaurum*, *M. mucogenicum*, *M. parafortuitum*, *M. smegmatis*, *M. chelonae*가 두 분석법에서 높은 score를 나타냈다. 16S rRNA 계통수에서도 *M. peregrinum*에 가까운 위치에 분포하였다. Figure 3.20, Figure 3.21는 *M. peregrinum*의 hsp65 염기 서열을 query로 한 blastn과 tblastx 결과를 종별로 추출한 것이다. *M. fortuitum*, *M. lentiflavum*, *M. wolinskyi*, *M. smegmatis*, *M. goodii* 등이 가장 score가 높은 종들이었다. Figure 3.20에서는 Figure 3.20에서는 *M. fortuitum*이 *M.*

peregrinum과 가장 유사한 반면, Figure 3.21에서 5종이 같은 score값으로 *M. peregrinum*과 유사성을 나타냈다. Figure 3.9의 hsp65 계통수에서도 *M. lentiflavum*을 제외한 나머지 종들은 진화적으로 *M. peregrinum*과 가장 가까운 종들이었다. Figure 3.22 와 Figure 3.23은 *M. peregrinum*의 rpoB 염기 서열을 query로 한 blastn과 tblastx 결과를 종별로 추출한 것이다. 두 분석 결과 공통적으로 *M. parafortuitum*이 *M. peregrinum*과 가장 유사한 종으로 나타났고 Figure 3.10의 rpoB 계통수에서도 확인할 수 있다. 그 외에 *M. fortuitum*, *M. phlei*, *M. smegmatis* 등이 두 분석에서 공통적으로 높은 score를 보였다. Figure 3.24 와 Figure 3.25은 *M. peregrinum*의 sodA 염기 서열을 query로 한 blastn과 tblastx 결과를 종별로 추출한 것이다. *M. mucogenicum*이 공통적으로 *M. fortuitum*과 가장 유사하였다. *M. wolinskyi*와 *M. goodii*가 공통적으로 높은 score를 나타내고 있다. 이는 Figure 3.11 sodA 계통수와 매우 유사한 결과임을 알 수 있다. Figure 3.26 와 Figure 3.27은 *M. peregrinum*의 tuf 염기 서열을 query로 한 blastn과 tblastx 결과를 종별로 추출한 것이다. 리스트 중에서 score가 높은 종이 score가 낮은 종보다 아래에 있는 경우는 e-value값이 증가하여 높은 유사성의 의미가 감소했기 때문이다. *M. fortuitum*이 공통적으로 가장 유사한 것으로 나타났지만 그 외에 높은 score를 나타낸 종들은 두 분석법에서 차이를 보였다. Tblastx 결과에서 *M. wolinskyi*, *M. goodii*, *M. smegmatis*, *M. nonchromogenicum*등이 높은 유사성을 보인 반면, blastn 결과에서는 *M. phocaicum*, *M. mucogenicum*, *M. aubagnense* 등으로 나타났다. 대부분의 분석 결과에서 *M. peregrinum*과 가장 유사한 종에 대해서 일치하였지만 그 이하 진화적으로 가까운 종들에 대해서는 조금씩 다른 양상을 보였다. 이처럼 같은 염기 서열을 query로 사용하

여도 blast 종류에 따라 다른 결과가 나오고 이는 계통수 분석과도 또 다른 접근법이기도 하다. 이와 같이 종 분류에 대한 연구는 어느 한 가지 분석법에 의존 할 수 없다. 더 많은 분석법이 개발될 필요가 있고 이를 종합하여 해석해야 함을 알 수 있다.

NTMDB
in LCBB



Welcome To NTMDB in LCBB

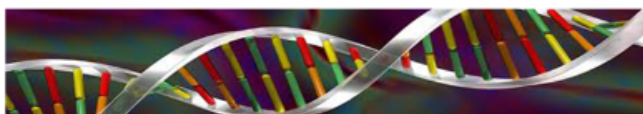
NTMDB is...

Well-organized database about NTM, by LCBB in SNU. NTMDB is a supportive database for the researchers who study for NTM and their marker genes. Using keyword in accession number, organism, gene, product users can search the gene and sequence data from 44 NTM species. Those data is also downloadable in FASTA form to be analyzed in ClustaX, Sequencematrix and MEGA6.0.

What is NTM?

The nontuberculosis mycobacteria refer to all the species in the family of mycobacteria that may cause human disease, but do not cause tuberculosis (TB).

Figure 3.1 The home page of NTMDB. The first page describes the database system and the definition of NTM.



[Home](#) [NTM gene search](#) [Sequence Fasta Files](#) [Marker-Sequence searching](#) [Blast&Analysis Programs](#)

[NTM Gene Search page]

This page provides the NTM gene table which the user can search with keywords; ACCESSION, ORGANISM, GENE, and PRODUCT. The table includes 5425 gene information of 44 NTM species. Using more than two keyword searching, the user can search the detailed gene information. The picture below shows the result in searching as an example.

example)

ACCESSION	ORGANISM	DEFINITION	CDS	GENE	PRODUCT
AM421322	Mycobacterium aubagnense	Mycobacterium aubagnense partial gyrB gene for DNA gyrase subunit B, type strain CCUG50186T.	<1..>314	gyrB	DNA gyrase subunit B

Check the boxes as many as you want to search, input the information, and push the "SEARCH" button below.

☐ ACCESSION

☒ ORGANISM

☒ GENE

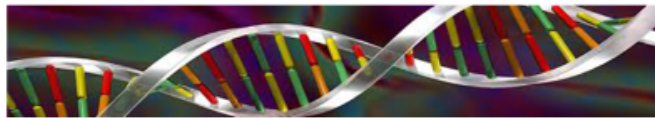
☐ PRODUCT

Find the NTM gene information you want!

select * from ntm where

ACCESSION	ORGANISM	DEFINITION	CDS	GENE	PRODUCT
-----------	----------	------------	-----	------	---------

Figure 3.2 “NTM Gene Search” page. This page supports the function of searching NTM gene information.



[Home](#) [NTM gene search](#) [Sequence Fasta Files](#) [Marker-Sequence searching](#) [Blast&Analysis Programs](#)

[NTM Gene Search page]

This page provides the NTM gene table which the user can search with keywords; ACCESSION, ORGANISM, GENE, and PRODUCT. The table includes 5425 gene information of 44 NTM species. Using more than two keyword searching, the user can search the detailed gene information. The picture below shows the result in searching as an example.

example)

ACCESSION	ORGANISM	DEFINITION	CDS	GENE	PRODUCT
AM421322	Mycobacterium aubagnense	Mycobacterium aubagnense partial gyrB gene for DNA gyrase subunit B, type strain CCUG50186T.	<1..>314	gyrB	DNA gyrase subunit B

Check the boxes as many as you want to search, input the information, and push the "SEARCH" button below.

☐ ACCESSION

☐ ORGANISM

☐ GENE

☐ PRODUCT

Find the NTM gene information you want!

select * from ntm where ORGANISM like '%Mycobacterium aubagnense%' and GENE like '%gyrB%'

ACCESSION	ORGANISM	DEFINITION	CDS	GENE	PRODUCT
AM421322	Mycobacterium aubagnense	Mycobacterium aubagnense partial gyrB gene for DNA gyrase subunit B, type strain CCUG50186T.	<1..>314	gyrB	DNA gyrase subunit B

Figure 3.3 The result of NTM Gene Searching. The result of searching the gyrB gene of *M. aubagnense* is shown up.



Figure 3.4 The function of “Marker-sequence Searching” page. The sequence of accession number, AM421322, is downloaded to the user.

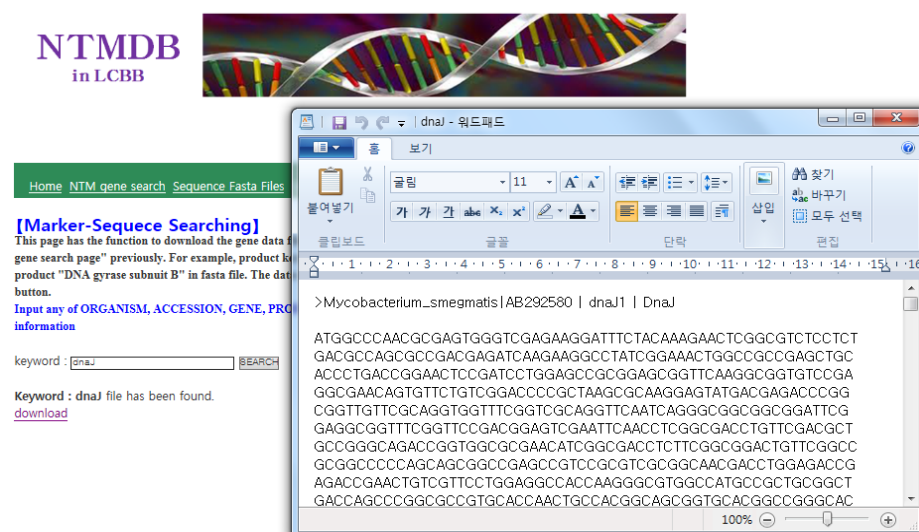


Figure 3.5 The result of searching other gene, dnaJ. “Marker-sequence searching” page supports the searching all the genes of NTM.

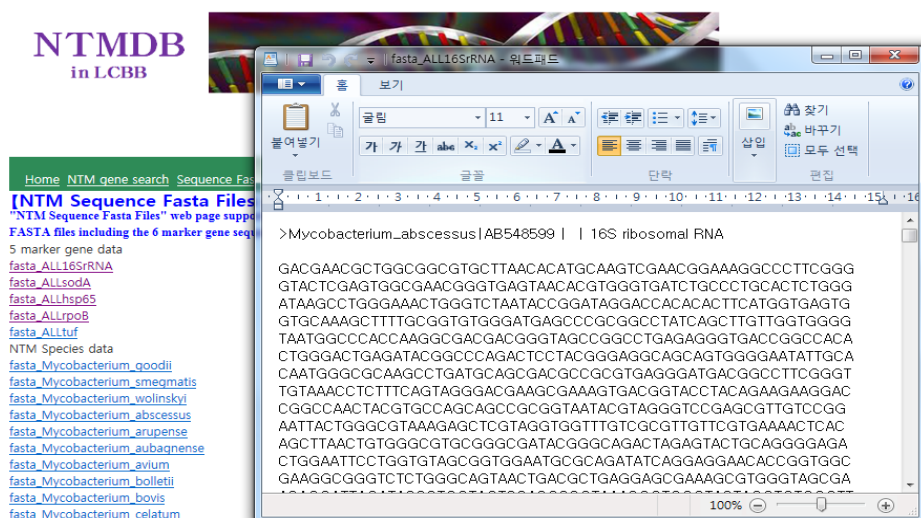


Figure 3.6 The function of “NTM Sequence Fasta Files” page. This page provides the sequence of 44 NTM species and 5 marker genes.

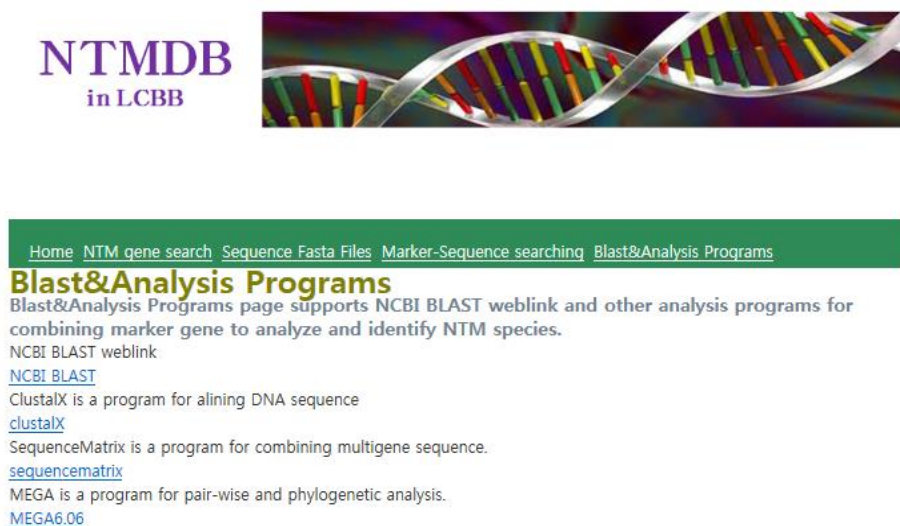


Figure 3.7 The function of “Blast & Analysis Programs” page. This page provides ClustalX, SequenceMatrix, MEGA programs, and NCBI BLAST web link.



Figure 3.8 Phylogenetic tree (NJ) of NTM using 16S rRNA gene. 33 NTM species are classified by 16S rRNA gene.

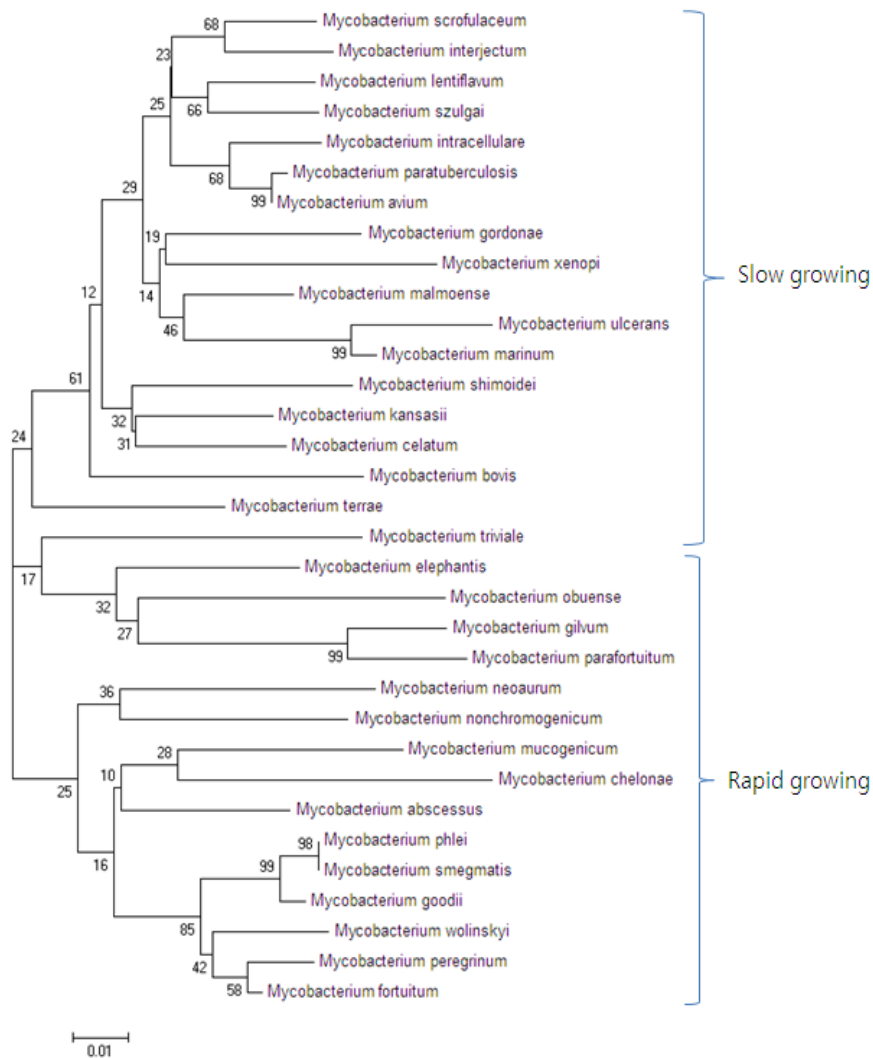


Figure 3.9 Phylogenetic tree (NJ) of NTM using hsp65 gene. 33 NTM species are classified by hsp65 gene.

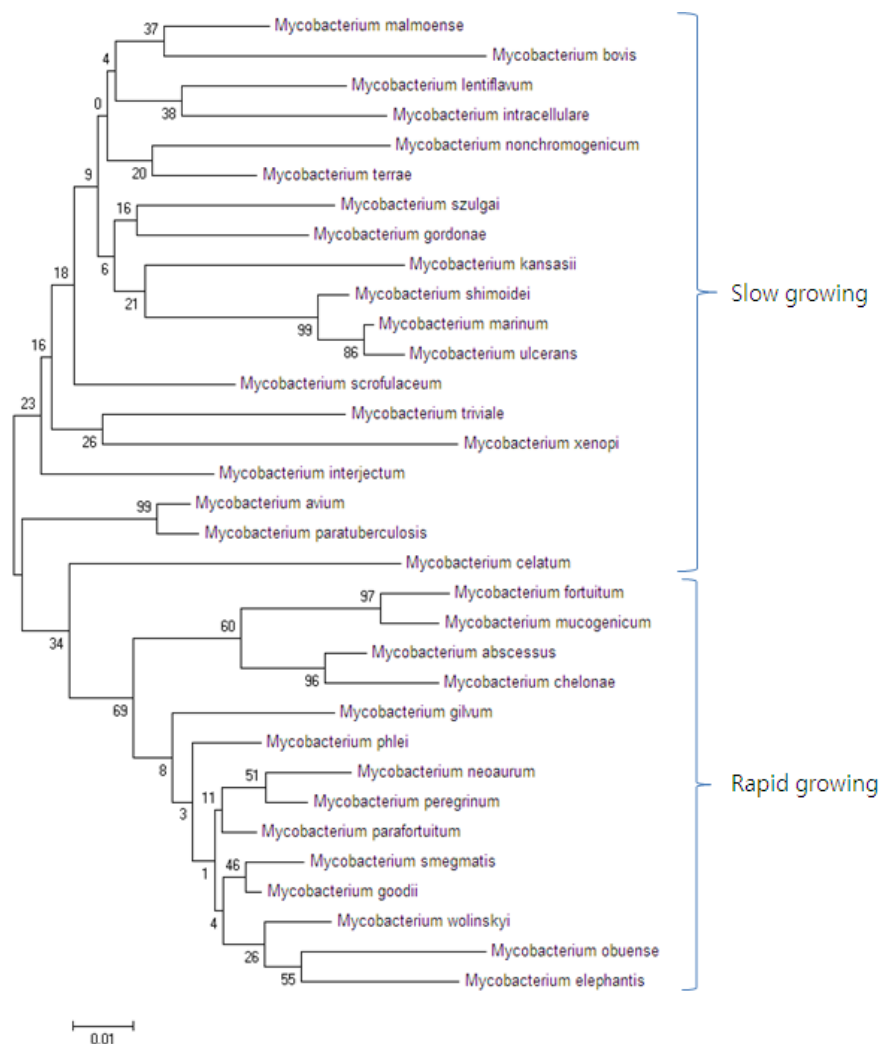


Figure 3.10 Phylogenetic tree (NJ) of NTM using *rpoB* gene. 33 NTM species are classified by *rpoB* gene.

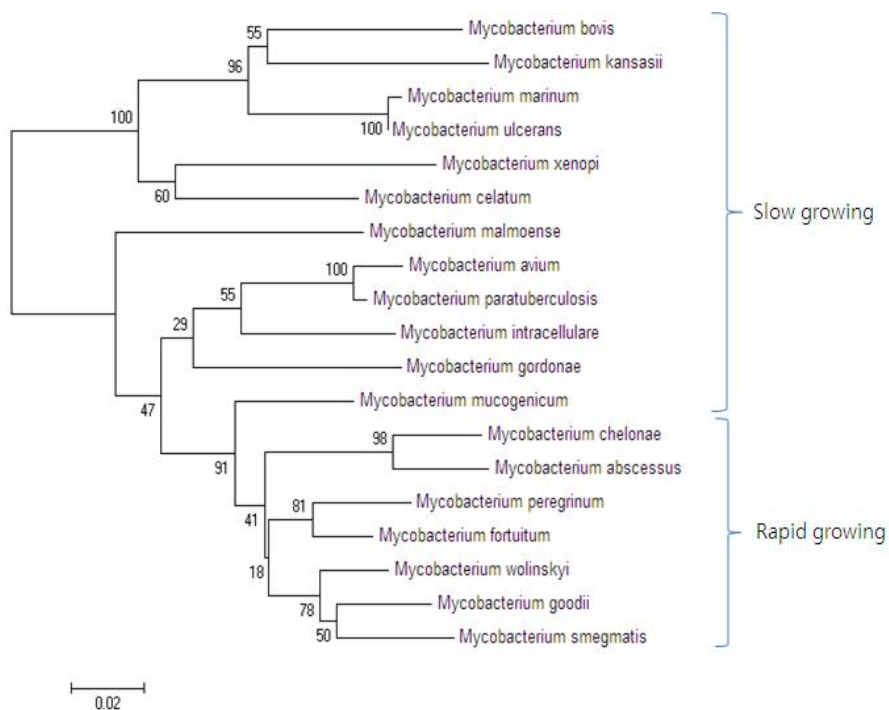


Figure 3.11 Phylogenetic tree (NJ) of NTM using *sodA* gene. 19 NTM species are classified by *sodA* gene.

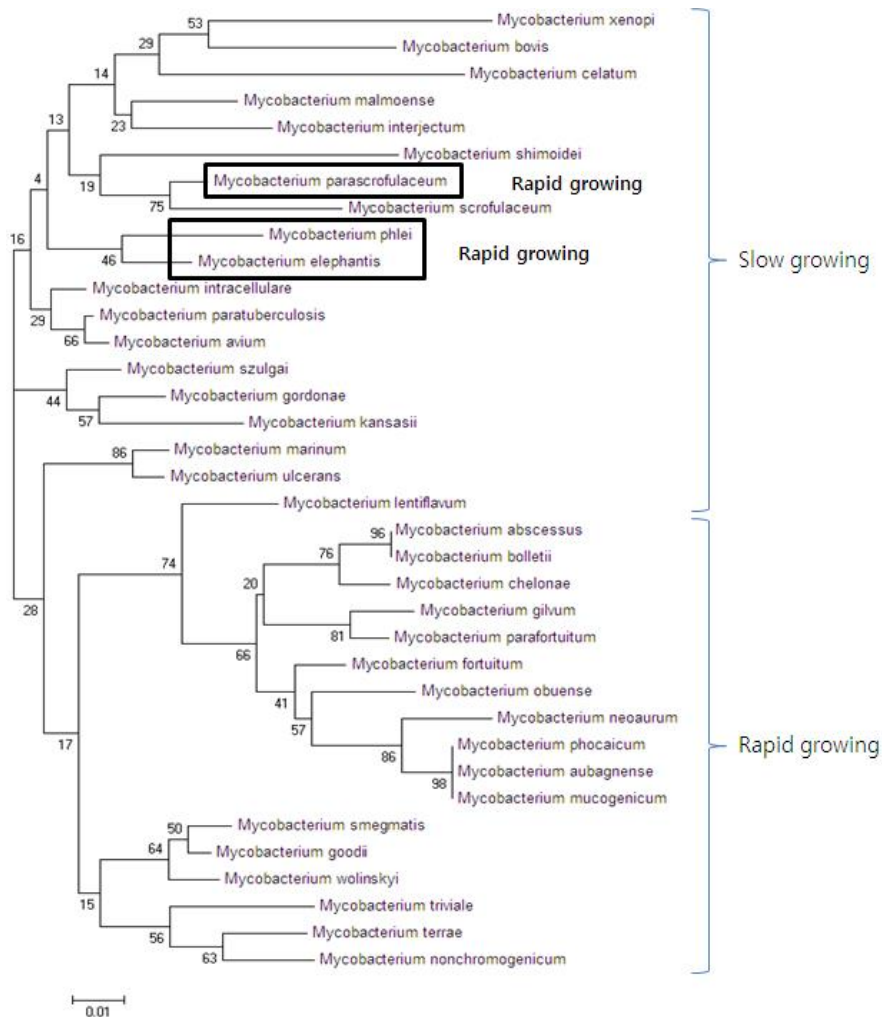


Figure 3.12 Phylogenetic tree (NJ) of NTM using *tuf* gene. 36 NTM species are classified by *tuf* gene.

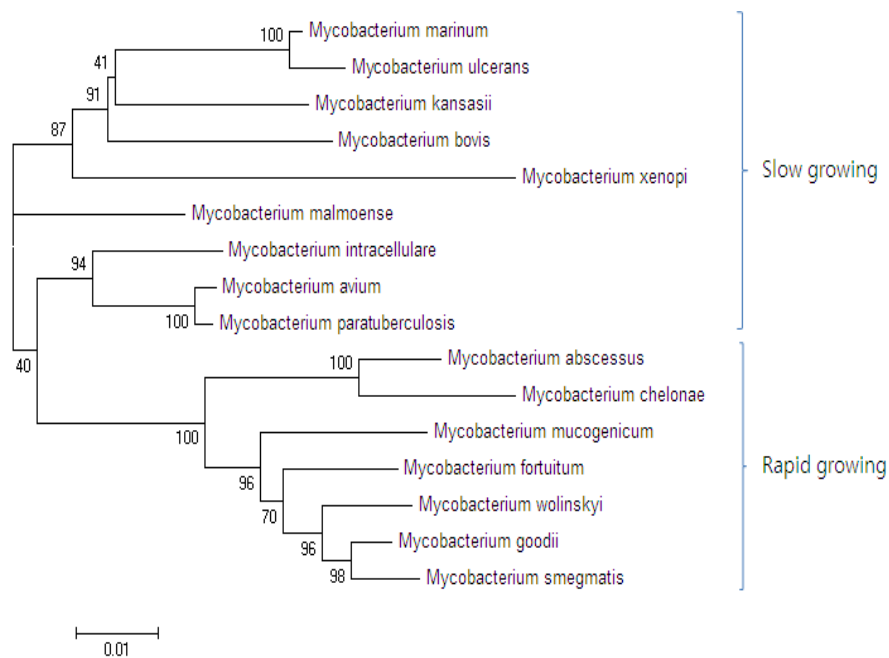


Figure 3.13 Phylogenetic tree (NJ) of NTM using 16S rRNA, hsp65, rpoB, sodA gene. 16 NTM species are classified by 16S rRNA, hsp65, rpoB, sodA gene.

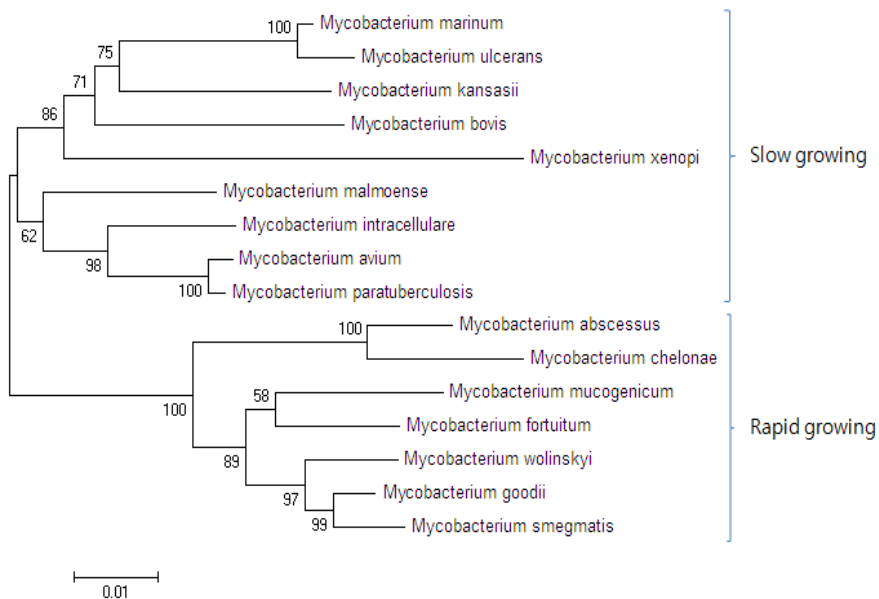


Figure 3.14 Phylogenetic tree (NJ) of NTM using 16S rRNA, hsp65, rpoB, sodA, tuf gene. 16 NTM species are classified by 16S rRNA, hsp65, rpoB, sodA, tuf gene.

a) 16S rRNA-hsp65-rpoB-sodA combination

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>Mycobacterium_abscessus</i>															
<i>Mycobacterium_avium</i>	0.068														
<i>Mycobacterium_bovis</i>	0.092	0.064													
<i>Mycobacterium_chelonae</i>	0.029	0.073	0.098												
<i>Mycobacterium_fortuitum</i>	0.050	0.065	0.083	0.061											
<i>Mycobacterium_goodii</i>	0.056	0.064	0.079	0.067	0.031										
<i>Mycobacterium_intracellulare</i>	0.072	0.031	0.058	0.076	0.069	0.070									
<i>Mycobacterium_kansasii</i>	0.084	0.058	0.051	0.096	0.083	0.079	0.061								
<i>Mycobacterium_malmoense</i>	0.075	0.042	0.058	0.081	0.073	0.067	0.040	0.053							
<i>Mycobacterium_marinum</i>	0.082	0.057	0.050	0.088	0.075	0.072	0.061	0.046	0.057						
<i>Mycobacterium_mucogenicum</i>	0.054	0.068	0.088	0.060	0.034	0.044	0.068	0.084	0.072	0.082					
<i>Mycobacterium_paratuberculosis</i>	0.066	0.005	0.064	0.071	0.064	0.062	0.031	0.058	0.043	0.059	0.067				
<i>Mycobacterium_smegmatis</i>	0.058	0.067	0.086	0.068	0.032	0.014	0.071	0.084	0.071	0.076	0.040	0.065			
<i>Mycobacterium_ulcerans</i>	0.087	0.064	0.054	0.091	0.080	0.076	0.065	0.050	0.061	0.009	0.087	0.065	0.080		
<i>Mycobacterium_wolinskyi</i>	0.051	0.066	0.085	0.069	0.028	0.020	0.072	0.082	0.073	0.076	0.040	0.065	0.024	0.083	
<i>Mycobacterium_xenopi</i>	0.104	0.083	0.082	0.110	0.092	0.087	0.088	0.080	0.080	0.079	0.100	0.084	0.094	0.083	0.091

b) 16S rRNA-hsp65-rpoB-sodA-tuf combination

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>Mycobacterium_abscessus</i>															
<i>Mycobacterium_avium</i>	0.071														
<i>Mycobacterium_bovis</i>	0.094	0.066													
<i>Mycobacterium_chelonae</i>	0.029	0.076	0.100												
<i>Mycobacterium_fortuitum</i>	0.048	0.068	0.084	0.059											
<i>Mycobacterium_goodii</i>	0.057	0.065	0.078	0.067	0.031										
<i>Mycobacterium_intracellulare</i>	0.073	0.030	0.060	0.077	0.070	0.069									
<i>Mycobacterium_kansasii</i>	0.086	0.058	0.056	0.097	0.086	0.080	0.061								
<i>Mycobacterium_malmoense</i>	0.078	0.042	0.059	0.084	0.076	0.067	0.041	0.053							
<i>Mycobacterium_marinum</i>	0.082	0.056	0.055	0.088	0.076	0.070	0.059	0.047	0.059						
<i>Mycobacterium_mucogenicum</i>	0.055	0.070	0.091	0.060	0.034	0.047	0.070	0.087	0.075	0.083					
<i>Mycobacterium_paratuberculosis</i>	0.069	0.005	0.065	0.074	0.066	0.062	0.029	0.058	0.043	0.057	0.069				
<i>Mycobacterium_smegmatis</i>	0.059	0.068	0.085	0.068	0.033	0.014	0.069	0.084	0.072	0.074	0.044	0.066			
<i>Mycobacterium_ulcerans</i>	0.087	0.061	0.058	0.092	0.081	0.075	0.062	0.051	0.062	0.009	0.087	0.062	0.079		
<i>Mycobacterium_wolinskyi</i>	0.053	0.066	0.085	0.068	0.030	0.020	0.070	0.082	0.074	0.074	0.045	0.065	0.023	0.080	
<i>Mycobacterium_xenopi</i>	0.106	0.085	0.082	0.112	0.095	0.087	0.088	0.083	0.082	0.082	0.103	0.085	0.095	0.085	0.091

Figure 3.15 P-distance Pairwise matrix using marker gene combination. Pairwise distance is calculated between each 16 species using a) 16S rRNA-hsp65-rpoB-sodA combination b) 16S rRNA-hsp65-rpoB-sodA-tuf combination

```
C:\Program Files\NCBI\blast-2.2.30\bin>makeblastdb -in db/16S_rRNA.fasta
-a -dbtype nucl -out db/16S_rRNA_thesisdb
```

```
Building a new DB, current time: 04/11/2015 21:33:55
New DB name:   db/16S_rRNA_thesisdb
New DB title:  db/16S_rRNA_thesisdb
Sequence type: Nucleotide
Deleted existing BLAST database with identical name.
Keep Linkouts: T
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1352 sequences in 0.199558 seconds.
```

```
C:\Program Files\NCBI\blast-2.2.30\bin>blastn -query db/16S_rRNA_peregrinum_only
.fasta -db db/16S_rRNA_thesisdb -out db/16S_rRNA_peregrinum_blastn_result.txt -num
_descriptions 1400
```

```
C:\Program Files\NCBI\blast-2.2.30\bin>tblastx -query db/16S_rRNA_peregrinum_onl
y.fasta -db db/16S_rRNA_thesisdb -out db/16S_rRNA_peregrinum_tblastx_result.txt -n
um_descriptions 1400
```

```
C:\Program Files\NCBI\blast-2.2.30\bin>_
```

Figure 3.16 WWWBlast database construction for 44 NTM species 5marker genes and blastn and tblastx execution. 5marker genes for 44 NTM nucleotide database constructed and 16S rRNA gene for *M. peregrinum* species used in blastn and tblastx.

BLASTN 2.2.30+

Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.

Database: db/fasta_ALL16SrRNA.fasta
1,352 sequences; 1,152,330 total letters

Query= Mycobacterium_peregrinum|AF058712 | | 16S ribosomal RNA

Length=1469

Sequences producing significant alignments:

	Score (Bits)	E Value
Mycobacterium_peregrinum AF058712 16S ribosomal RNA	2710	0.0
Mycobacterium_wolinskyi AY457083 16S ribosomal RNA	2636	0.0
Mycobacterium_smegmatis_str._MC2_155 CP001663 rrsA 16S ribo...	2610	0.0
Mycobacterium_smegmatis_str._MC2_155 CP001663 rrsB 16S ribo...	2610	0.0
Mycobacterium_smegmatis_str._MC2_155 CP000480 rrsB 16S ribo...	2610	0.0
Mycobacterium_smegmatis_str._MC2_155 CP000480 rrsA 16S ribo...	2610	0.0
Mycobacterium_peregrinum AF537362 16S ribosomal RNA	2595	0.0
Mycobacterium_smegmatis_str._MC2_155 Y08453 16S rRNA 16S ri...	2591	0.0
Mycobacterium_smegmatis AJ131761 16S rRNA 16S ribosomal RNA	2591	0.0
Mycobacterium_gilvum_Spyr1 CP002385 16S ribosomal RNA	2588	0.0
Mycobacterium_gilvum_Spyr1 CP002385 16S ribosomal RNA	2588	0.0
Mycobacterium_gilvum_PVR-GCK CP000656 AAPA01000000 AAPA01000001...	2588	0.0
Mycobacterium_gilvum_PVR-GCK CP000656 AAPA01000000 AAPA01000001...	2588	0.0

> Mycobacterium_peregrinum|AF058712 | | 16S ribosomal RNA

Length=1469

Score = 2710 bits (1467), Expect = 0.0
Identities = 1469/1469 (100%), Gaps = 0/1469 (0%)
Strand=Plus/Plus

Query	1	ACACATGCAAGTCGAACGGAAAGGCCCTTCGGGGTGCTCGAGTGGCGAACGGGTGAGTAA	60
Sbjct	1	ACACATGCAAGTCGAACGGAAAGGCCCTTCGGGGTGCTCGAGTGGCGAACGGGTGAGTAA	60
Query	61	CACGTGGGTGATCTGCCCTGCACCTTTGGGATAAGCCTGGGAACTGGGTCTAATACCGAA	120
Sbjct	61	CACGTGGGTGATCTGCCCTGCACCTTTGGGATAAGCCTGGGAACTGGGTCTAATACCGAA	120
Query	121	TAGGACTACGGTCTGCATGG#CTGTGGTGGAAAGCTTTTGCGGTGTGGGATGGGCCCGCG	180
Sbjct	121	TAGGACTACGGTCTGCATGG#CTGTGGTGGAAAGCTTTTGCGGTGTGGGATGGGCCCGCG	180
Query	181	GCCTATCAGCTTGTGGTGGGGTGATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGA	240
Sbjct	181	GCCTATCAGCTTGTGGTGGGGTGATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGA	240
Query	241	GAGGGTGACDGGCCACACTGGGACTGAGATACGGCCAGACTCCTACGGGAGGCAGCAGT	300
Sbjct	241	GAGGGTGACDGGCCACACTGGGACTGAGATACGGCCAGACTCCTACGGGAGGCAGCAGT	300
Query	301	GGGGAATATTGCACAAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGC	360
Sbjct	301	GGGGAATATTGCACAAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGC	360

Figure 3.17 The result of blastn for *M. peregrinum* 16S rRNA gene. The query is, gene id = AF058712. The bit score and identity analysis is displayed.

Mycobacterium_peregrinum AF058712 16S ribosomal RNA	2710	0.0
Mycobacterium_wolinskyi AV457083 16S ribosomal RNA	2636	0.0
Mycobacterium_smegmatis_str._MC2_155 CP001663 rrsA 16S ribo...	2610	0.0
Mycobacterium_gilvum_Spyr1 CP002385 16S ribosomal RNA	2588	0.0
Mycobacterium_neoaurum_VKM_Ac-1815D CP006936 ANB101000000 ANB10...	2577	0.0
Mycobacterium_goodii AV457079 16S ribosomal RNA	2577	0.0
Mycobacterium_fortuitum_subsp._acetamidolyticum FR733720 16S ...	2575	0.0
Mycobacterium_parafortuitum X93183 16S rRNA 16S ribosomal RNA	2536	0.0
Mycobacterium_mucogenicum AM884317 16S rRNA 16S ribosomal RNA	2497	0.0
Mycobacterium_elephantis AJ536100 16S rRNA 16S ribosomal RNA	2494	0.0
Mycobacterium_chelonae JX010972 16S ribosomal RNA	2488	0.0
Mycobacterium_phocaicum AY859682 16S ribosomal RNA	2463	0.0
Mycobacterium_phlei_DSM_43239_=_CCUG_21000 AF480603 16S ri...	2464	0.0
Mycobacterium_aubagnense AY859683 16S ribosomal RNA	2464	0.0
Mycobacterium_obuense JQ660308 16S ribosomal RNA	2451	0.0
Mycobacterium_kubicae HM022200 16S-23S ribosomal RNA inter...	2436	0.0
Mycobacterium_terrae HM770865 16S ribosomal RNA	2420	0.0
Mycobacterium_gordonae AY604571 internal transcribed spacer	2418	0.0
Mycobacterium_marinum_E11 HG917972 rrs 16S ribosomal RNA	2407	0.0
Mycobacterium_lentiflavum AB362380 16S ribosomal RNA	2403	0.0
Mycobacterium_shimoidei AJ005005 16S rRNA 16S ribosomal RNA	2401	0.0
Mycobacterium_ulcerans_Agy99 CP000325 rrs Ribosomal RNA 16S	2394	0.0
Mycobacterium_arupense AM884329 16S rRNA 16S ribosomal RNA	2392	0.0
Mycobacterium_abscessus_subsp._bolletii_str._GO_06 CP003699 ...	2390	0.0
Mycobacterium_abscessus DQ666776 16S-23S ribosomal RNA int...	2390	0.0
Mycobacterium_szulgai JX629457 16S ribosomal RNA	2385	0.0
Mycobacterium_seoulense DQ536403 16S ribosomal RNA	2383	0.0
Mycobacterium_paraseoulense DQ536404 16S ribosomal RNA	2383	0.0
Mycobacterium_senuense DQ536408 16S ribosomal RNA	2381	0.0
Mycobacterium_bovis_BCG_str._ATCC_35743 CP003494 AEZH01000000 A...	2381	0.0
Mycobacterium_triviale X88924 16S ribosomal RNA	2372	0.0
Mycobacterium_parascrofulaceum AF152559 16S ribosomal RNA	2370	0.0
Mycobacterium_malmoense AF367025 internal transcribed spacer	2368	0.0
Mycobacterium_scrofulaceum X52924 16S ribosomal RNA	2364	0.0
Mycobacterium_interjectum X70961 16S ribosomal RNA	2364	0.0
Mycobacterium_avium_subsp._hominissuis_TH135 AP012555 16S ...	2364	0.0
Mycobacterium_nonchromogenicum X52928 16S ribosomal RNA	2359	0.0
Mycobacterium_intracellulare_MOTT-64 CP003324 16S ribosoma...	2359	0.0
Mycobacterium_shinjukuense AB268503 16S ribosomal RNA	2348	0.0
Mycobacterium_kansasii_ATCC_12478 CP006835 ACBV01000000 ACBV010...	2346	0.0
Mycobacterium_celatum L08169 16S ribosomal RNA	2338	0.0
Mycobacterium_avium_subsp._paratuberculosis_MAP4 CP005928 rrs...	2309	0.0
Mycobacterium_timonense JX976611 16S ribosomal RNA	2289	0.0
Mycobacterium_xenopi AJ536033 16S rRNA 16S ribosomal RNA	2224	0.0

Figure 3.18 The result of 16S rRNA blastn. Parsing each NTM species from 16S rRNA blastn result by scoring order.

<i>Mycobacterium_peregrinum</i> AF058712 16S ribosomal RNA	1125	0.0	1
<i>Mycobacterium_wolinskyi</i> AY457083 16S ribosomal RNA	1095	0.0	1
<i>Mycobacterium_gilvum</i> _Spyr1 CP002385 16S ribosomal RNA	1090	0.0	1
<i>Mycobacterium_neoaurum</i> _VKM_Ac-18150 CP006936 ANB101000000 ANB10...	1080	0.0	1
<i>Mycobacterium_fortuitum</i> EU741213 16S ribosomal RNA	567	0.0	3
<i>Mycobacterium_mucogenicum</i> AM884317 16S rRNA 16S ribosomal RNA	590	0.0	3
<i>Mycobacterium_chelonae</i> JX010972 16S ribosomal RNA	593	0.0	3
<i>Mycobacterium_phocaicum</i> AY859682 16S ribosomal RNA	590	0.0	3
<i>Mycobacterium_paraafortuitum</i> X93183 16S rRNA 16S ribosomal RNA	1071	0.0	1
<i>Mycobacterium_smegmatis</i> _str._MC2_155 CP001663 rrsA 16S ribo...	516	0.0	5
<i>Mycobacterium_gordonae</i> X581472 16S rRNA 16S ribosomal RNA	733	0.0	3
<i>Mycobacterium_aubagnense</i> AY859683 16S ribosomal RNA	586	0.0	3
<i>Mycobacterium_obuense</i> JQ660308 16S ribosomal RNA	732	0.0	3
<i>Mycobacterium_elephantis</i> AJ010747 16S rRNA 16S ribosomal RNA	741	0.0	5
<i>Mycobacterium_szulgai</i> JX629457 16S ribosomal RNA	726	0.0	4
<i>Mycobacterium_goodii</i> AY457079 16S ribosomal RNA	989	0.0	3
<i>Mycobacterium_marinum</i> AY509246 16S ribosomal RNA	606	0.0	5
<i>Mycobacterium_shimoidei</i> AJ005005 16S rRNA 16S ribosomal RNA	336	0.0	6
<i>Mycobacterium_malmoense</i> AF367025 internal transcribed spacer	726	0.0	4
<i>Mycobacterium_shinjukuense</i> AB268503 16S ribosomal RNA	720	0.0	4
<i>Mycobacterium_avium</i> _subsp._hominissuis_TH135 AP012555 16S ...	721	0.0	4
<i>Mycobacterium_ulcerans</i> _Agg99 CP000325 rrs Ribosomal RNA 16S	606	0.0	5
<i>Mycobacterium_terrae</i> HM770865 16S ribosomal RNA	349	0.0	6
<i>Mycobacterium_intracellulare</i> _MOTT-64 CP003324 16S ribosoma...	716	0.0	4
<i>Mycobacterium_kansasii</i> _ATCC_12478 CP006835 ACBVO1000000 ACBVO10...	722	0.0	4
<i>Mycobacterium_kubicae</i> HM022200 16S-23S ribosomal RNA inter...	494	0.0	4
<i>Mycobacterium_bovis</i> M20940 16S ribosomal RNA	608	0.0	6
<i>Mycobacterium_scrofulaceum</i> X52924 16S ribosomal RNA	715	0.0	4
<i>Mycobacterium_sensuense</i> DQ536408 16S ribosomal RNA	338	0.0	6
<i>Mycobacterium_celatum</i> L08170 16S ribosomal RNA	609	0.0	5
<i>Mycobacterium_seoulense</i> DQ536403 16S ribosomal RNA	722	0.0	5
<i>Mycobacterium_paraseoulense</i> DQ536404 16S ribosomal RNA	722	0.0	5
<i>Mycobacterium_arupense</i> AM884329 16S rRNA 16S ribosomal RNA	343	0.0	6
<i>Mycobacterium_phlei</i> _DSM_43239_=_CCUG_21000 AF480603 16S ri...	968	0.0	2
<i>Mycobacterium_avium</i> _subsp._paratuberculosis_MAP4 CP005928 rrs...	721	0.0	4
<i>Mycobacterium_abscessus</i> _subsp._bolletii_str._GO_06 CP003699 ...	855	0.0	3
<i>Mycobacterium_abscessus</i> DQ866776 16S-23S ribosomal RNA int...	854	0.0	3
<i>Mycobacterium_timonense</i> JX976611 16S ribosomal RNA	717	0.0	4
<i>Mycobacterium_lentiflavum</i> AB362380 16S ribosomal RNA	496	0.0	4
<i>Mycobacterium_nonchromogenicum</i> X52928 16S ribosomal RNA	340	0.0	6
<i>Mycobacterium_triviale</i> X88924 16S ribosomal RNA	499	0.0	3
<i>Mycobacterium_parascrofulaceum</i> AF152559 16S ribosomal RNA	487	0.0	4
<i>Mycobacterium_xenopi</i> AJ536033 16S rRNA 16S ribosomal RNA	708	0.0	5
<i>Mycobacterium_interjectum</i> X70961 16S ribosomal RNA	498	0.0	3

Figure 3.19 The result of 16S rRNA tblastx. Parsing each NTM species from 16S rRNA tblastx result by scoring order.

<i>Mycobacterium peregrinum</i> AY458069 hsp65 65 kDa heat shock p...	741	0.0
<i>Mycobacterium fortuitum</i> AF140677 65 kDa heat shock protein	691	0.0
<i>Mycobacterium lentii</i> Favum JX294391 hsp65 heat shock protein...	689	0.0
<i>Mycobacterium wolinskyi</i> FJ531484 hsp65 65 kDa heat shock pr...	675	0.0
<i>Mycobacterium smegmatis</i> HM454230 hsp65 65 kDa heat shock pr...	647	0.0
<i>Mycobacterium goodii</i> AY458071 hsp65 65 kDa heat shock protein	647	0.0
<i>Mycobacterium abscessus</i> JX154121 hsp65 65 kDa heat shock pr...	608	9e-176
<i>Mycobacterium mucogenicum</i> EF551421 hsp65 65 kDa heat shock ...	608	9e-176
<i>Mycobacterium phocaicum</i> EU266579 hsp65 65 kDa heat shock pr...	603	4e-174
<i>Mycobacterium abscessus</i> _subsp._ <i>boletii</i> JX154114 hsp65 65 k...	592	9e-171
<i>Mycobacterium phlei</i> U17952 65 kDa heat shock protein	582	6e-168
<i>Mycobacterium neoaurum</i> FJ515915 hsp65 65 kDa heat shock pro...	580	2e-167
<i>Mycobacterium obuense</i> AF547863 hsp65 65 kDa heat shock protein	575	1e-165
<i>Mycobacterium aubagnense</i> DQ987727 hsp65 65 kDa heat shock p...	569	4e-164
<i>Mycobacterium terrae</i> JX154097 hsp65 65 kDa heat shock protein	562	7e-162
<i>Mycobacterium nonchromogenicum</i> JX294382 hsp65 heat shock pr...	562	7e-162
<i>Mycobacterium parafortuitum</i> AF547864 hsp65 65 kDa heat shoc...	558	1e-160
<i>Mycobacterium sensuense</i> JN571200 hsp65 65 kDa heat shock pro...	555	1e-159
<i>Mycobacterium kubicae</i> AF547850 hsp65 65 kDa heat shock protein	553	4e-159
<i>Mycobacterium interjectum</i> AY550236 hsp65 65 kDa heat shock ...	547	2e-157
<i>Mycobacterium avium</i> _complex HM056132 hsp65 65 kDa heat shock...	542	1e-155
<i>Mycobacterium xenopi</i> AY550207 hsp65 65 kDa heat shock protein	536	5e-154
<i>Mycobacterium shinjukuense</i> AB268505 hsp65 65 kDa heat shock...	536	5e-154
<i>Mycobacterium scrofulaceum</i> AY550227 hsp65 65 kDa heat shock...	536	5e-154
<i>Mycobacterium avium</i> _subsp._ <i>paratuberculosis</i> X74518 HSP65 he...	536	5e-154
<i>Mycobacterium parascrofulaceum</i> AY337276 hsp65 65 kDa heat s...	536	5e-154
<i>Mycobacterium gordonae</i> EU486081 hsp65 65 kDa heat shock pro...	536	5e-154
<i>Mycobacterium gilvum</i> AF547838 hsp65 65 kDa heat shock protein	536	5e-154
<i>Mycobacterium celatum</i> JQ582663 65 kDa heat shock protein	536	5e-154
<i>Mycobacterium chelonae</i> U17928 65 kDa heat shock protein	532	6e-153
<i>Mycobacterium marinum</i> HM359134 hsp65 65 kDa heat shock protein	525	1e-150
<i>Mycobacterium kansasii</i> KJ186613 hsp65 65 kDa heat shock pro...	525	1e-150
<i>Mycobacterium elephantis</i> AF547828 hsp65 65 kDa heat shock p...	525	1e-150
<i>Mycobacterium mageritense</i> HQ263568 hsp65 65 kDa heat shock pr...	520	5e-149
<i>Mycobacterium intracellulare</i> KF030699 65 kDa heat shock pr...	520	5e-149
<i>Mycobacterium arupense</i> FJ263631 hsp65 65 kDa heat shock pro...	514	2e-147
<i>Mycobacterium ulcerans</i> AY550209 hsp65 65 kDa heat shock pro...	514	2e-147
<i>Mycobacterium triviale</i> AF547883 hsp65 65 kDa heat shock pro...	514	2e-147
<i>Mycobacterium timonense</i> KJ364654 hsp65 65 kDa heat shock pr...	503	5e-144
<i>Mycobacterium szulgai</i> KC481265 hsp65 heat shock protein 65 kDa	503	5e-144
<i>Mycobacterium shimoidei</i> AY299140 hsp65 heat shock protein 65	484	2e-138

Figure 3.20 The result of hsp65 blastn. Parsing each NTM species from hsp65 blastn result by scoring order.

<i>Mycobacterium_smegmatis</i> AJ307653 hsp65 65 kDa heat shock pr...	191	2e-061	2
<i>Mycobacterium_peregrinum</i> AM398455 hsp65 65 kDa heat shock p...	191	2e-061	2
<i>Mycobacterium_fortuitum</i> AM398469 hsp65 65 kDa heat shock pr...	191	2e-061	2
<i>Mycobacterium_lentiflavum</i> JX294391 hsp65 heat shock protein...	191	3e-061	2
<i>Mycobacterium_wolinskyi</i> AY458064 hsp65 65 kDa heat shock pr...	191	3e-061	2
<i>Mycobacterium_goodii</i> AY458071 hsp65 65 kDa heat shock protein	191	3e-061	2
<i>Mycobacterium_abscessus</i> E0732719 hsp65 heat shock protein 65	189	6e-061	2
<i>Mycobacterium_abscessus_subsp._bolletii</i> HQ404297 hsp65 heat...	189	6e-061	2
<i>Mycobacterium_szulgai</i> FJ384764 hsp65 65 kDa heat shock protein	188	2e-060	2
<i>Mycobacterium_ulcerans_Agy99</i> AB548728 hsp65 heat shock prot...	187	4e-060	2
<i>Mycobacterium_marinum</i> DQ066751 hsp65 heat shock protein 65	187	4e-060	2
<i>Mycobacterium_intracellulare</i> KF186675 65 kDa heat shock pr...	187	4e-060	2
<i>Mycobacterium_interjectum</i> AY550236 hsp65 65 kDa heat shock ...	187	4e-060	2
<i>Mycobacterium_avium</i> AY550214 hsp65 65 kDa heat shock protein	187	4e-060	2
<i>Mycobacterium_timonense</i> KJ364654 hsp65 65 kDa heat shock pr...	187	4e-060	2
<i>Mycobacterium_avium_subsp._paratuberculosis</i> AF547809 hsp65 ...	187	4e-060	2
<i>Mycobacterium_kubicae</i> AF547850 hsp65 65 kDa heat shock protein	187	4e-060	2
<i>Mycobacterium_shinjukuense</i> AB268505 hsp65 65 kDa heat shock...	186	5e-060	2
<i>Mycobacterium_scrofulaceum</i> AY550228 hsp65 65 kDa heat shock...	186	5e-060	2
<i>Mycobacterium_bovis</i> AY438082 hsp65 65 kDa heat shock protein	186	5e-060	2
<i>Mycobacterium_parascrofulaceum</i> HM454227 hsp65 65 kDa heat s...	186	6e-060	2
<i>Mycobacterium_kansasii</i> KJ186613 hsp65 65 kDa heat shock pro...	185	1e-059	2
<i>Mycobacterium_celatum</i> JQ582663 65 kDa heat shock protein	185	1e-059	2
<i>Mycobacterium_nonchromogenicum</i> JN571193 hsp65 65 kDa heat s...	185	2e-059	2
<i>Mycobacterium_gordonae</i> JN964684 hsp65 heat shock protein 65	185	2e-059	2
<i>Mycobacterium_phocaicum</i> KF667480 hsp65 heat shock protein 6...	185	2e-059	2
<i>Mycobacterium_malmoense</i> HQ263569 hsp65 65 kDa heat shock pr...	184	2e-059	2
<i>Mycobacterium_terrae</i> JN571204 hsp65 65 kDa heat shock protein	183	4e-059	2
<i>Mycobacterium_paraseoulense</i> HM602042 hsp65 65 kDa heat shoc...	183	6e-059	2
<i>Mycobacterium_xenopi</i> AY550207 hsp65 65 kDa heat shock protein	182	1e-058	2
<i>Mycobacterium_shimoiidei</i> AF547874 hsp65 65 kDa heat shock pr...	182	1e-058	2
<i>Mycobacterium_mucogenicum</i> AM398473 hsp65 65 kDa heat shock ...	181	3e-058	2
<i>Mycobacterium_chelonae</i> AM398467 hsp65 65 kDa heat shock pro...	178	1e-057	2
<i>Mycobacterium_aubagnense</i> DQ987727 hsp65 65 kDa heat shock p...	178	2e-057	2
<i>Mycobacterium_neoaurum_ATCC_25795</i> AF547860 hsp65 65 kDa hea...	178	2e-057	2
<i>Mycobacterium_sensuense</i> JN571200 hsp65 65 kDa heat shock pro...	178	2e-057	2
<i>Mycobacterium_phlei</i> JX294389 hsp65 heat shock protein 65 kDa	177	3e-057	2
<i>Mycobacterium_obuense</i> AF547863 hsp65 65 kDa heat shock protein	177	3e-057	2
<i>Mycobacterium_elephantis</i> AF547828 hsp65 65 kDa heat shock p...	177	3e-057	2
<i>Mycobacterium_arupense</i> JN571190 hsp65 65 kDa heat shock pro...	175	1e-056	2
<i>Mycobacterium_parafortuitum</i> AF547864 hsp65 65 kDa heat shoc...	175	1e-056	2
<i>Mycobacterium_gilvum</i> AF547838 hsp65 65 kDa heat shock protein	174	2e-056	2
<i>Mycobacterium_triviale</i> AF434737 hsp65 65 kDa heat shock pro...	173	4e-056	2
<i>Mycobacterium_seoulense</i> DQ536401 heat shock protein 65	183	3e-055	2

Figure 3.21 The result of hsp65 tblastx. Parsing each NTM species from hsp65 tblastx result by scoring order.

<i>Mycobacterium_peregrinum</i> AY147166 rpoB RpoB	566	6e-163
<i>Mycobacterium_parafortuitum</i> EF472355 rpoB RNA polymerase beta	521	1e-149
<i>Mycobacterium_fortuitum</i> AY147172 rpoB RpoB	518	2e-148
<i>Mycobacterium_smegmatis_str._MC2_155</i> CP000480 rpoB DNA-dire...	516	6e-148
<i>Mycobacterium_neoaurum</i> AF057477 rpoB RNA polymerase beta	510	3e-146
<i>Mycobacterium_phlei_DSM_43239</i> =_CCUG_21000 AF057480 rpoB RN...	505	1e-144
<i>Mycobacterium_goodii</i> AY544918 rpoB RNA polymerase	501	2e-143
<i>Mycobacterium_golinskyi</i> AY544978 rpoB RNA polymerase	496	7e-142
<i>Mycobacterium_chelonae</i> EU732715 rpoB RNA polymerase beta su...	492	1e-140
<i>Mycobacterium_abscessus_subsp._bolletii</i> KF360857 rpoB DNA-d...	492	1e-140
<i>Mycobacterium_abscessus</i> AY147164 rpoB RpoB	475	1e-135
<i>Mycobacterium_mucogenicum</i> AY147171 rpoB RpoB	473	3e-135
<i>Mycobacterium_aubagnense</i> AY859694 rpoB RNA polymerase subun...	473	3e-135
<i>Mycobacterium_phocaicum</i> AY859693 rpoB RNA polymerase subunit B	468	2e-133
<i>Mycobacterium_obuense</i> AY544951 rpoB RNA polymerase	459	1e-130
<i>Mycobacterium_gilvum</i> AY544917 rpoB RNA polymerase	457	4e-130
<i>Mycobacterium_interjectum</i> AF057470 rpoB RNA polymerase beta	444	3e-126
<i>Mycobacterium_kubicae</i> AY544937 rpoB RNA polymerase	440	4e-125
<i>Mycobacterium_elephantis</i> AY544908 rpoB RNA polymerase	435	2e-123
<i>Mycobacterium_avium</i> KJ499915 rpoB DNA-directed RNA polymera...	431	2e-122
<i>Mycobacterium_szulgai</i> KC481264 rpoB RNA polymerase beta-sub...	425	1e-120
<i>Mycobacterium_avium_subsp._paratuberculosis</i> AF057479 rpoB R...	425	1e-120
<i>Mycobacterium_parascrofulaceum</i> JQ411538 rpoB RNA polymerase...	425	1e-120
<i>Mycobacterium_triviale</i> JN037842 rpoB RNA polymerase beta su...	422	1e-119
<i>Mycobacterium_gordonae</i> FJ384766 rpoB RNA polymerase	420	5e-119
<i>Mycobacterium_shimoidei</i> AY544962 rpoB RNA polymerase	418	2e-118
<i>Mycobacterium_terrae</i> AF057488 rpoB RNA polymerase beta	414	2e-117
<i>Mycobacterium_scrofulaceum</i> AF057482 rpoB RNA polymerase beta	414	2e-117
<i>Mycobacterium_intracellulare</i> KJ499914 rpoB DNA-directed RNA...	414	2e-117
<i>Mycobacterium_timonense</i> HQ692916 rpoB RNA polymerase B	409	1e-115
<i>Mycobacterium_nonchromogenicum</i> AF057478 rpoB RNA polymerase...	409	1e-115
<i>Mycobacterium_malmoense</i> AF057475 rpoB RNA polymerase beta	403	5e-114
<i>Mycobacterium_sensuense</i> FJ268578 rpoB RNA polymerase	392	1e-110
<i>Mycobacterium_marinum</i> E11 HG917972 rpoB DNA-directed RNA po...	392	1e-110
<i>Mycobacterium_kansasii</i> AF173086 rpoB RNA polymerase beta su...	392	1e-110
<i>Mycobacterium_xenopi</i> AF057493 rpoB RNA polymerase beta	388	1e-109
<i>Mycobacterium_arupense</i> AB239923 rpoB RNA polymerase beta su...	388	1e-109
<i>Mycobacterium_ulcerans</i> AB624294 rpoB RNA polymerase	387	5e-109
<i>Mycobacterium_paraseoulense</i> HQ287519 rpoB RNA polymerase B	387	5e-109
<i>Mycobacterium_lentiflavum</i> AB362383 rpoB RNA polymerase	387	5e-109
<i>Mycobacterium_celatum</i> JQ599258 rpoB RNA polymerase	385	2e-108
<i>Mycobacterium_shinjukuense</i> HG673735 rpoB RNA polymerase	370	5e-104
<i>Mycobacterium_seoulense</i> EU191926 rpoB RNA polymerase	366	6e-103
<i>Mycobacterium_bovis_BCG_str._Mexico</i> CP002095 rpoB DNA-direc...	364	2e-102

Figure 3.22 The result of rpoB blastn. Parsing each NTM species from rpoB blastn result by scoring order.

Mycobacterium_parafortuitum EF472959 rpoB RNA polymerase beta	254	4e-070	1
Mycobacterium_phiferi_DSM_45235_-CCUG_21000 AF057480 rpoB RNA...	251	5e-069	1
Mycobacterium_smeagmatis_str._MC2_155 CP000480 rpoB DNA-dire...	248	3e-068	1
Mycobacterium_wolinskyi AY544978 rpoB RNA polymerase	247	7e-068	1
Mycobacterium_fortuitum AY147172 rpoB RpoB	246	1e-067	1
Mycobacterium_goodii AY544918 rpoB RNA polymerase	246	2e-067	1
Mycobacterium_neoaurum AF057477 rpoB RNA polymerase beta	243	8e-067	1
Mycobacterium_aubagnense AY859694 rpoB RNA polymerase subun...	243	2e-066	1
Mycobacterium_triviale JN037842 rpoB RNA polymerase beta su...	242	3e-066	1
Mycobacterium_kubicae AY544937 rpoB RNA polymerase	241	4e-066	1
Mycobacterium_interjectum AF057470 rpoB RNA polymerase beta	241	6e-066	1
Mycobacterium_nonchromogenicum AF057478 rpoB RNA polymerase...	240	8e-066	1
Mycobacterium_gilvum AY544917 rpoB RNA polymerase	240	8e-066	1
Mycobacterium_shimoidei AY544962 rpoB RNA polymerase	235	3e-064	1
Mycobacterium_abscessus_subsp._boletii KF360857 rpoB DNA-d...	235	3e-064	1
Mycobacterium_terrae AF057488 rpoB RNA polymerase beta	235	3e-064	1
Mycobacterium_gordonae FJ384766 rpoB RNA polymerase	235	3e-064	1
Mycobacterium_ulcerans_Ayy99 CP000325 rpoB DNA-directed RNA...	234	5e-064	1
Mycobacterium_sensuense FJ268578 rpoB RNA polymerase	234	5e-064	1
Mycobacterium_marinum_E11 HG917972 rpoB DNA-directed RNA po...	234	5e-064	1
Mycobacterium_chelonae EU732715 rpoB RNA polymerase beta su...	234	5e-064	1
Mycobacterium_szulgai KC481264 rpoB RNA polymerase beta-sub...	233	1e-063	1
Mycobacterium_phocaicum AY859693 rpoB RNA polymerase subunit B	232	2e-063	1
Mycobacterium_parascrofulaceum JQ411538 rpoB RNA polymerase...	232	2e-063	1
Mycobacterium_mucogenicum DQ369730 rpoB RNA polymerase	232	2e-063	1
Mycobacterium_elephantis AY544908 rpoB RNA polymerase	232	2e-063	1
Mycobacterium_arupense AB239923 rpoB RNA polymerase beta su...	232	2e-063	1
Mycobacterium_xenopi AF057493 rpoB RNA polymerase beta	232	3e-063	1
Mycobacterium_malmoense AF057475 rpoB RNA polymerase beta	231	4e-063	1
Mycobacterium_celatum AF057459 rpoB RNA polymerase beta	231	4e-063	1
Mycobacterium_avium KJ499916 rpoB DNA-directed RNA polymera...	231	4e-063	1
Mycobacterium_abscessus AY544879 rpoB RNA polymerase	231	4e-063	1
Mycobacterium_obuense AY544951 rpoB RNA polymerase	230	8e-063	1
Mycobacterium_avium_subsp._paratuberculosis AF057479 rpoB R...	230	1e-062	1
Mycobacterium_timonense HQ692916 rpoB RNA polymerase B	229	2e-062	1
Mycobacterium_shinjuense HG673735 rpoB RNA polymerase	229	2e-062	1
Mycobacterium_scrofulaceum AF057482 rpoB RNA polymerase beta	229	2e-062	1
Mycobacterium_pareseoulense HQ267519 rpoB RNA polymerase B	229	2e-062	1
Mycobacterium_kansasii AF173088 rpoB RNA polymerase beta su...	229	2e-062	1
Mycobacterium_intracellulare KJ499914 rpoB DNA-directed RNA...	229	2e-062	1
Mycobacterium_bovis_BCG_str._Mexico CP002095 rpoB DNA-direc...	229	2e-062	1
Mycobacterium_lentiflavum AB362383 rpoB RNA polymerase	227	6e-062	1
Mycobacterium_seoulense EU191926 rpoB RNA polymerase	216	2e-058	1

Figure 3.23 The result of rpoB tblastx. Parsing each NTM species from rpoB tblastx result by scoring order.

Mycobacterium_peregrinum AY458115 sodA superoxide dismutase	815	0.0
Mycobacterium_fortuitum_subsp._fortuitum_DSM_46621 AY458114 s...	704	0.0
Mycobacterium_phocaicum AY859706 sodA SodA	627	0.0
Mycobacterium_wolinskyi AY458119 sodA superoxide dismutase	621	0.0
Mycobacterium_mucogenicum AY458111 sodA superoxide dismutase	621	0.0
Mycobacterium_goodii AY458107 sodA superoxide dismutase	599	4e-174
Mycobacterium_smegmatis AY458120 sodA superoxide dismutase	593	2e-172
Mycobacterium_aubagnense AY859707 sodA SodA	582	4e-169
Mycobacterium_gordonae HQ263579 sodA superoxide dismutase	556	2e-161
Mycobacterium_abscessus AY458102 sodA superoxide dismutase	544	2e-157
Mycobacterium_abscessus_subsp._bolletii AY593975 sodA SodA	538	9e-156
Mycobacterium_chelonae AY458105 sodA superoxide dismutase	521	9e-151
Mycobacterium_malmoense HQ263572 sodA superoxide dismutase	520	3e-150
Mycobacterium_intracellulare_ATCC_13950 JQ411518 sodA super...	505	9e-146
Mycobacterium_avium_subsp._hominissuis EU409984 sodA supero...	505	9e-146
Mycobacterium_avium_subsp._paratuberculosis EU409988 sodA s...	499	4e-144
Mycobacterium_arupense EU191932 sodA superoxide dismutase	418	1e-119

Figure 3.24 The result of *sodA* blastn. Parsing each NTM species from *sodA* blastn result by scoring order.

Mycobacterium_peregrinum AY458115 sodA superoxide dismutase	365	2e-104	1
Mycobacterium_fortuitum_subsp._fortuitum_DSM_46621 AY458114 s...	365	2e-104	1
Mycobacterium_wolinskyi AY458119 sodA superoxide dismutase	358	3e-102	1
Mycobacterium_smegmatis AY458120 sodA superoxide dismutase	355	2e-101	1
Mycobacterium_goodii AY458107 sodA superoxide dismutase	352	1e-100	1
Mycobacterium_abscessus_subsp._bolletii AY862403 sodA SodA	352	1e-100	1
Mycobacterium_abscessus AY458102 sodA superoxide dismutase	352	1e-100	1
Mycobacterium_mucogenicum AY458110 sodA superoxide dismutase	349	9e-100	1
Mycobacterium_aubagnense AY859707 sodA SodA	349	9e-100	1
Mycobacterium_chelonae AY458105 sodA superoxide dismutase	348	2e-099	1
Mycobacterium_phocaicum AY859706 sodA SodA	347	5e-099	1
Mycobacterium_avium_subsp._paratuberculosis EU409988 sodA s...	339	1e-096	1
Mycobacterium_avium_subsp._hominissuis EU409990 sodA supero...	339	1e-096	1
Mycobacterium_intracellulare_ATCC_13950 JQ411518 sodA super...	338	2e-096	1
Mycobacterium_malmoense HQ263574 sodA superoxide dismutase	338	3e-096	1
Mycobacterium_gordonae HQ263579 sodA superoxide dismutase	335	2e-095	1
Mycobacterium_bovis_BCG_str._Mexico CP002095 sodA Superoxid...	296	1e-083	1
Mycobacterium_ulcerans_Agy99 CP000325 sodA superoxide dismu...	294	5e-083	1
Mycobacterium_parascrofulaceum JQ411517 sodA superoxide dis...	294	5e-083	1
Mycobacterium_marinum_E11 HG917972 sodA superoxide dismutas...	294	5e-083	1
Mycobacterium_celatum AB824286 sodA superoxide dismutase	291	4e-082	1
Mycobacterium_arupense EU191932 sodA superoxide dismutase	287	4e-081	1
Mycobacterium_xenopi HQ263581 sodA superoxide dismutase	281	3e-079	1
Mycobacterium_kansasii EU127837 sodA superoxide dismutase	246	1e-068	1

Figure 3.25 The result of *sodA* tblastx. Parsing each NTM species from *sodA* tblastx result by scoring order.

<i>Mycobacterium peregrinum</i> DQ986085 tuf elongation factor Tu	1199	0.0
<i>Mycobacterium fortuitum</i> _subsp._ <i>fortuitum</i> _DSM_46621 DQ986087 t...	1144	0.0
<i>Mycobacterium phocaicum</i> DQ986084 tuf elongation factor Tu	989	0.0
<i>Mycobacterium mucogenicum</i> DQ986082 tuf elongation factor Tu	977	0.0
<i>Mycobacterium aubagnense</i> DQ986083 tuf elongation factor Tu	977	0.0
<i>Mycobacterium sensuense</i> HQ450860 tuf elongation factor Tu	939	0.0
<i>Mycobacterium obuense</i> DQ986109 tuf elongation factor Tu	939	0.0
<i>Mycobacterium smegmatis</i> DQ986129 tuf elongation factor Tu	928	0.0
<i>Mycobacterium neoaurum</i> _ATCC_25795 DQ986095 tuf elongation f...	922	0.0
<i>Mycobacterium parafortuitum</i> DQ986113 tuf elongation factor Tu	917	0.0
<i>Mycobacterium goodii</i> DQ986127 tuf elongation factor Tu	917	0.0
<i>Mycobacterium nonchromogenicum</i> DQ986064 tuf elongation fact...	911	0.0
<i>Mycobacterium wolinskyi</i> DQ986118 tuf elongation factor Tu	900	0.0
<i>Mycobacterium parascrofulaceum</i> DQ986124 tuf elongation fact...	900	0.0
<i>Mycobacterium intracellulare</i> _MOTT-64 CP003324 elongation f...	900	0.0
<i>Mycobacterium gilvum</i> DQ986112 tuf elongation factor Tu	900	0.0
<i>Mycobacterium lentiflavum</i> DQ986028 tuf elongation factor Tu	894	0.0
<i>Mycobacterium arupense</i> EU191944 tuf elongation factor Tu	891	0.0
<i>Mycobacterium triviale</i> DQ986068 tuf elongation factor Tu	889	0.0
<i>Mycobacterium avium</i> _subsp._ <i>paratuberculosis</i> DQ986049 tuf el...	883	0.0
<i>Mycobacterium avium</i> _subsp._ <i>hominissuis</i> _TH135 AP012555 tuf e...	883	0.0
<i>Mycobacterium timonense</i> HQ450853 tuf elongation factor Tu	878	0.0
<i>Mycobacterium terrae</i> DQ986066 tuf elongation factor Tu	878	0.0
<i>Mycobacterium elephantis</i> DQ986116 tuf elongation factor Tu	872	0.0
<i>Mycobacterium abscessus</i> _subsp._ <i>bolletii</i> _50594 CP004374 elo...	863	0.0
<i>Mycobacterium kansasii</i> DQ986039 tuf elongation factor Tu	861	0.0
<i>Mycobacterium seoulense</i> EU191942 tuf elongation factor Tu	857	0.0
<i>Mycobacterium phlei</i> DQ986106 tuf elongation factor Tu	857	0.0
<i>Mycobacterium abscessus</i> DQ986060 tuf elongation factor Tu	857	0.0
<i>Mycobacterium interjectum</i> DQ986057 tuf elongation factor Tu	856	0.0
<i>Mycobacterium kubicae</i> DQ986017 tuf elongation factor Tu	852	0.0
<i>Mycobacterium chelonae</i> DQ986059 tuf elongation factor Tu	846	0.0
<i>Mycobacterium szulgai</i> DQ986011 tuf elongation factor Tu	845	0.0
<i>Mycobacterium scrofulaceum</i> DQ986046 tuf elongation factor Tu	845	0.0
<i>Mycobacterium paraseoulense</i> HQ287514 tuf elongation factor Tu	839	0.0
<i>Mycobacterium gordonae</i> DQ986010 tuf elongation factor Tu	817	0.0
<i>Mycobacterium shimoidei</i> DQ986026 tuf elongation factor Tu	811	0.0
<i>Mycobacterium ulcerans</i> DQ986016 tuf elongation factor Tu	800	0.0
<i>Mycobacterium marinum</i> DQ986013 tuf elongation factor Tu	800	0.0
<i>Mycobacterium celatum</i> DQ986042 tuf elongation factor Tu	797	0.0
<i>Mycobacterium bovis</i> _BCG_str._ <i>Mexico</i> JN034606 tuf elongation...	795	0.0
<i>Mycobacterium malmoense</i> DQ986055 tuf elongation factor Tu	789	0.0
<i>Mycobacterium xenopi</i> DQ986009 tuf elongation factor Tu	763	0.0
<i>Mycobacterium shinjukuense</i> HG673741 tuf elongation factor Tu	745	0.0

Figure 3.26 The result of tuf blastn. Parsing each NTM species from tuf blastn result by scoring order.

<i>Mycobacterium peregrinum</i> DQ986085 tuf elongation factor Tu	142	8e-100	4
<i>Mycobacterium fortuitum</i> _subsp._ <i>fortuitum</i> _DSM_46621 DQ986087 t...	142	6e-098	4
<i>Mycobacterium wolinskyi</i> DQ986118 tuf elongation factor Tu	134	2e-091	4
<i>Mycobacterium goodii</i> DQ986127 tuf elongation factor Tu	136	1e-090	4
<i>Mycobacterium smegmatis</i> _str._MC2_155 CP001663 tuf Elongatio...	134	3e-090	4
<i>Mycobacterium kubicai</i> DQ986017 tuf elongation factor Tu	129	4e-090	4
<i>Mycobacterium nonchromogenicum</i> DQ986064 tuf elongation fact...	133	5e-090	4
<i>Mycobacterium arupense</i> EU191944 tuf elongation factor Tu	133	9e-090	4
<i>Mycobacterium obuense</i> DQ986109 tuf elongation factor Tu	134	1e-089	3
<i>Mycobacterium sensuense</i> HQ450860 tuf elongation factor Tu	134	3e-089	4
<i>Mycobacterium ulcerans</i> DQ986016 tuf elongation factor Tu	135	4e-089	3
<i>Mycobacterium marinum</i> DQ986013 tuf elongation factor Tu	135	4e-089	3
<i>Mycobacterium elephantis</i> DQ986116 tuf elongation factor Tu	137	6e-089	3
<i>Mycobacterium lentiflavum</i> DQ986028 tuf elongation factor Tu	134	6e-089	3
<i>Mycobacterium kansasii</i> _ATCC_12478 CP006835 ACBV01000000 ACBV010...	136	8e-089	3
<i>Mycobacterium phocaicum</i> DQ986084 tuf elongation factor Tu	132	8e-089	3
<i>Mycobacterium mucogenicum</i> DQ986082 tuf elongation factor Tu	132	8e-089	3
<i>Mycobacterium aubagnense</i> DQ986083 tuf elongation factor Tu	132	8e-089	3
<i>Mycobacterium avium</i> _subsp._ <i>hominissuis</i> _TH135 AP012555 tuf e...	134	2e-088	4
<i>Mycobacterium avium</i> _subsp._ <i>paratuberculosis</i> DQ986049 tuf el...	134	2e-088	4
<i>Mycobacterium phlei</i> DQ986106 tuf elongation factor Tu	136	2e-088	3
<i>Mycobacterium timonense</i> HQ450853 tuf elongation factor Tu	135	3e-088	3
<i>Mycobacterium parafortuitum</i> DQ986113 tuf elongation factor Tu	132	4e-088	3
<i>Mycobacterium gilvum</i> DQ986112 tuf elongation factor Tu	132	4e-088	3
<i>Mycobacterium neoaurum</i> _VKM_Ac-18150 CP006936 ANB101000000 ANB10...	132	7e-088	3
<i>Mycobacterium paraseoulense</i> HQ287514 tuf elongation factor Tu	132	7e-088	3
<i>Mycobacterium shimoidae</i> DQ986026 tuf elongation factor Tu	134	9e-088	3
<i>Mycobacterium szulgai</i> DQ986011 tuf elongation factor Tu	132	9e-088	3
<i>Mycobacterium shinjukuense</i> HG573741 tuf elongation factor Tu	133	1e-087	3
<i>Mycobacterium interjectum</i> DQ986057 tuf elongation factor Tu	133	1e-087	3
<i>Mycobacterium parascrofulaceum</i> DQ986124 tuf elongation fact...	132	2e-087	3
<i>Mycobacterium triviale</i> DQ986068 tuf elongation factor Tu	132	2e-087	3
<i>Mycobacterium scrofulaceum</i> DQ986046 tuf elongation factor Tu	132	2e-087	3
<i>Mycobacterium intracellulare</i> _MOTT-64 CP003324 elongation f...	132	2e-087	3
<i>Mycobacterium gordonae</i> DQ986010 tuf elongation factor Tu	130	2e-087	3
<i>Mycobacterium bovis</i> _BCG_str._Mexico JN034606 tuf elongation...	131	4e-087	3
<i>Mycobacterium seoulense</i> EU191942 tuf elongation factor Tu	129	4e-087	3
<i>Mycobacterium celatum</i> DQ986042 tuf elongation factor Tu	131	6e-087	3
<i>Mycobacterium terrae</i> DQ986066 tuf elongation factor Tu	136	8e-087	3
<i>Mycobacterium malmoense</i> DQ986055 tuf elongation factor Tu	130	1e-086	3
<i>Mycobacterium xenopi</i> DQ986009 tuf elongation factor Tu	131	9e-085	3
<i>Mycobacterium abscessus</i> _subsp._ <i>bolletii</i> _50594 CP004374 elo...	132	1e-083	4
<i>Mycobacterium abscessus</i> CU458896 Elongation factor Tu (EF-Tu)	132	1e-083	4
<i>Mycobacterium chelonae</i> DQ986059 tuf elongation factor Tu	132	1e-082	3

Figure 3.27 The result of tuf tblastx. Parsing each NTM species from tuf tblastx result by scoring order.

Table 3.1 Marker gene data status

Category \ Marker gene	16S rRNA	hsp65	rpoB	sodA	tuf
Number of NTM species	44	44	44	24	44
Total number of sequence	1352	828	841	55	69

5 marker gene status is described.

Table 3.2 NTM species data status

Species	Total number of sequence	Marker gene status	Species	Total number of sequence	Marker gene status
M. goodii	26	16S rRNA, hsp65, rpoB, sodA, tuf	M. nonchromogenicum	53	16S rRNA, hsp65, rpoB, tuf
M. smegmatis	20806	16S rRNA, hsp65, rpoB, sodA, tuf	M. obuense	16	16S rRNA, hsp65, rpoB, tuf
M. wolinskyi	16	16S rRNA, hsp65, rpoB, sodA, tuf	M. paraafortuitum	18	16S rRNA, hsp65, rpoB, tuf
M. abscessus	5411	16S rRNA, hsp65, rpoB, sodA, tuf	M. parascrofulaceum	58	16S rRNA, hsp65, rpoB, sodA, tuf
M. arupense	79	16S rRNA, hsp65, rpoB, sodA, tuf	M. paraseoulense	9	16S rRNA, hsp65, rpoB, tuf
M. aubagnense	14	16S rRNA, hsp65, rpoB, sodA, tuf	M. paratuberculosis	9034	16S rRNA, hsp65, rpoB, sodA, tuf
M. avium	10871	16S rRNA, hsp65, rpoB, sodA, tuf	M. peregrinum	78	16S rRNA, hsp65, rpoB, sodA, tuf
M. bolletii	10599	16S rRNA, hsp65, rpoB, sodA, tuf	M. phlei	47	16S rRNA, hsp65, rpoB, tuf
M. celatum	80	16S rRNA, hsp65, rpoB, sodA, tuf	M. phocaicum	49	16S rRNA, hsp65, rpoB, sodA, tuf
M. chelonae	244	16S rRNA, hsp65, rpoB,	M. scrofulaceum	63	16S rRNA, hsp65, rpoB,

		sodA, tuf			tuf
M. elephantis	17	16S rRNA, hsp65, rpoB, tuf	M. senuense	17	16S rRNA, hsp65, rpoB, tuf
M. fortuitum	266	16S rRNA, hsp65, rpoB, sodA, tuf	M. seoulense	11	16S rRNA, hsp65, rpoB, tuf
M. gilvum	11029	16S rRNA, hsp65, rpoB, tuf	M. shimoidei	26	16S rRNA, hsp65, rpoB, tuf
M. gordonae	246	16S rRNA, hsp65, rpoB, sodA, tuf	M. szulgai	63	16S rRNA, hsp65, rpoB, tuf
M. interjectum	32	16S rRNA, hsp65, rpoB, tuf	M. terrae	57	16S rRNA, hsp65, rpoB, tuf
M. kansasii	6106	16S rRNA, hsp65, rpoB, sodA, tuf	M. timonense	17	16S rRNA, hsp65, rpoB, tuf
M. kubicae	11	16S rRNA, hsp65, rpoB, tuf	M. triviale	40	16S rRNA, hsp65, rpoB, tuf
M. lentiflavum	41	16S rRNA, hsp65, rpoB, tuf	M. ulcerans	4412	16S rRNA, hsp65, rpoB, sodA, tuf
M. malmoense	50	16S rRNA, hsp65, rpoB, sodA, tuf	M. xenopi	84	16S rRNA, hsp65, rpoB, sodA, tuf
M. marinum	11222	16S rRNA, hsp65, rpoB, sodA, tuf	M. shinjukuense	6	16S rRNA, hsp65, rpoB, tuf
M. mucogenicum	121	16S rRNA, hsp65, rpoB, sodA, tuf	M. bovis	28461	16S rRNA, hsp65, rpoB, sodA, tuf
M. neoaurum	5210	16S rRNA, hsp65, rpoB, tuf	M. intracellulare	15979	16S rRNA, hsp65, rpoB, sodA, tuf

44 species data status is described.

Table 3.3 The results of marker gene analysis

Gene	Average fragment size (bp)	species included (n)	Mean p-distance	Mean K2P distance	G+C content (mol%)	Bootstrap $\geq 50\%$	Bootstrap $\geq 75\%$
hsp65	413.9	33	0.0098	0.106	65.8	36.67	20.00
rpoB	334.2	33	0.088	0.094	66.9	30.00	16.67
sodA	464.3	19	0.146	0.165	63.4	75.00	50.00
16S rRNA	957.2	33	0.024	0.024	57.4	56.25	26.67
tuf	163.1	36	0.087	0.094	65.2	53.13	21.88
4markers	2183.9	16	0.066	0.070	61.5	84.62	76.92
5markers	2346.9	16	0.067	0.071	61.8	100.00	76.92

5 marker genes and two combination analysis results are described (4markers: hsp65, rpoB, sodA, 16S rRNA; 5markers: hsp65, rpoB, sodA, 16S rRNA, tuf).

Table 3.4 The result values of WWWBlast for *M. peregrinum*

Category	BLAST	blastn	tblastx
	lambda(/gapped)	1.33/1.28	0.318
	K(/gapped)	0.621/0.460	0.134
the number of 16S rRNA sequence output		1077	1115
the number of hsp65 sequence output		692	811
the number of rpoB sequence output		340	405
the number of sodA sequence output		38	55
the number of tuf sequence output		67	69
	matrix	blastn matrix 1 -2	BLOSUM62

The comparable results of blastn and tblastx are described. The values after forward slashes refer to the gapped variables of blastn.

제 4 장. 고 찰

4.1 고 찰

본 연구는 기존에 없는 NTM에 대한 유전 정보와 서열에 대한 오픈 소스의 데이터베이스를 구축하여 NTM에 대한 다양한 연구가 가능하도록 하였다는 것에 의의가 있다. 또한 마커 유전자의 다양한 조합을 통한 다각적인 생명정보학적인 분석은 NTM 고유의 동정법에 소요되는 시간과 비용을 상당 부분 절감시킬 것으로 기대한다. 생명정보학적 NTM균 동정법은 연구자뿐 아니라 환자의 조기 진단 및 치료가 가능해지고 치료 비용도 절감할 수 있을 것이다. 본 데이터베이스를 통한 NTM 연구는 기존 연구에 비해 NTM 종간에 분류학적 구분을 명확히 하여 보건학적으로 각 종에 대한 역학적 접근이 가능할 것이다. 이를 통해 최근 몇 년 간 증가하고 있는 NTM 질병과 관련된 종에 대한 세분화된 연구가 가능하고 각 종이 일으키는 질병과 위험 요인에 대해 분석할 수 있다.

본 연구의 기대성과를 세부적으로 살펴보면, 학술적으로는 기존의 실험적이고 형태학적인 NTM 분석법의 한계점을 극복하고 생물학적 이론과 첨단 컴퓨터 기술이 결합한 생명정보학적 분석을 통해 NTM 연구에 접목시킴으로써 새로운 연구 분야의 길을 제시할 수 있다는 것이며, NTM과 더불어 다양한 생물과 질병에 관한 연구에서 생명정보학적 연구가 이루어 질 수 있을 것으로 기대된다. 기술적 측면에서는 생명정보학적 분석법은 기존의 분석법과 다르게

다각도로 무궁무진한 접근이 가능하고 인간으로서는 불가능한 대용량 정보 처리를 컴퓨터를 이용하여 실행되기 때문에 컴퓨터 기술의 발전과 함께 앞으로도 무한한 발전 가능성을 제시할 것이다. 기존의 분석법은 결과를 어느 정도 기대하고 시행하는 분석법인 것에 비해, 생명정보학적 분석법은 다양한 분석 기법을 다양한 방법으로 조합하여 실행 가능하기 때문에 이를 통해 축적된 수많은 분석 결과는 또 다른 분석에 쓰여서 학술적으로 의미 있는 분석 성과를 기하급수적으로 도출할 것으로 기대된다. 산업적인 측면에서는 의학, 보건학, 약학과 같이 건강과 직결된 산업 분야뿐 아니라 의류업, 식품업, 군수업, 농수산업 같이 일반 생활과 밀접한 연관이 있는 전반적인 산업 분야에 걸쳐 NTM과 같은 박테리아균의 유전적 정보는 크게 영향을 미칠 것으로 기대된다. 세균의 접근을 막는 신소재를 이용한 의류 산업의 발전, 인간에게 유익한 균을 배양하는 식품의 개발, 세균을 이용하여 화학 무기에 피해를 입은 군인들을 긴급 치료할 수 있는 의약품의 개발, 토양과 농작물에 유익한 균의 배양 등 다양한 산업 분야에서 이용될 수 있는 유전적 정보를 제공할 수 있는 토대를 마련할 수 있을 것이다. 본 연구는 특히, 보건학적인 측면에서 활용도가 매우 클 것으로 기대된다. NTM의 빠른 종 동정은 진단 기간을 단축하여 신속한 환자의 치료로 이어져 환자의 건강을 지키는 보건학적인 역할을 수행할 수 있다. 또한 다양한 생화학적 절차를 상당히 축소할 수 있기 때문에 그에 필요한 많은 인력과 경비를 절약하고 정확하고 효율적인 생명정보학적 동종 검사가 가능하다. 다양한 연구자에 의한 꾸준한 업데이트를 통해 현재 환자에게서 많이 분리되고 있는 NTM 종에 대한 역학 연구가 가능하고 이를 통해 NTM 질환 발병의 환경적 요인에 대한 빠른 예방 조치를 취할 수

있다. 본 데이터베이스의 NTM종의 유전자 서열 검색을 통해 특정 유전자를 타겟으로 하는 의약품 개발에도 크게 이바지 할 것으로 기대된다. 최근에 많이 연구되고 있는 유전자 타겟 의약품은 특정 유전자의 전사(transcription)나 단백질 활성화(activity)를 방해함으로써 해당 균이 발병을 일으키지 못하도록 하는 것으로 유전자 서열과 단백질의 대사활동의 파악이 매우 중요하다. 그런 측면에서 본 데이터베이스는 NTM의 특정 유전자 서열 검색과 다운로드를 통해 유전자 타겟 의약품이 작용할 수 있는 유전자 파악이 가능하고 이를 연구에 활용이 가능하다.

본 연구는 웹 서버를 기반으로 오픈 소스의 NTM 데이터베이스를 구축함으로써 사용하는 NTM 종이 보유하고 있는 유전자 검색과 특정 유전자를 지닌 종의 검색이 모두 가능하여 사용자가 필요한 종에 대한 유전자 데이터를 다운 받고 분석에 활용할 수 있다. 특히 최근에 많이 이루어지고 있는 다양한 마커 유전자를 조합하여 계통 분류 분석이 가능하게 하여 NTM 종 동정이 가능하도록 하였다. NTM은 단일의 생화학적, 임상학적 검사로 결핵과 구분이 어렵기 때문에 다각도의 동정법이 필요하다. 마커 유전자도 단일 유전자 서열 비교로는 NTM 종간의 분류 분석이 정확한 결과를 얻기 어렵다. 현재까지의 다양한 마커 유전자 조합을 이용한 많은 연구들도 결과가 조금씩 다르기 때문에 더 많은 마커 유전자의 발견과 정확한 분류 동정을 위한 마커 유전자들 간의 다양한 조합을 시도할 필요가 있다. NTM 동정을 위한 마커 유전자는 많은 종이 보유하고 있고 생명 활동에 필수적인 기능을 수행하는 유전자(housekeeping gene)이어야 한다. 그리고 서열 정보를 이용한 동정에서 가까운 종 간에 분별력이 커야 마커 유전자로서의 활용도가 높다(Shin et al., 2008 ;

Adekambi et al., 2004). 본 연구 분석에 활용된 5가지 마커 유전자는 많은 NTM 종들이 공통적으로 보유하고 박테리아 유전자의 전사(transcription), 번역(translation) 과정, 면역 반응 등에 관여하는 필수적인 기능을 나타내었다. 앞으로 마커로 활용될 유전자는 NMT종의 유전자와 기능에 대한 연구를 통해 검증할 수 있다. 잠재적 마커 유전자는 NTM 동정법에 활용하여 종 간의 분별력이 기존 마커 유전자들보다 우수함이 밝혀지면 마커 유전자로 활용할 수 있다. 본 데이터베이스를 이용하면 사용자가 원하는 유전 서열 정보 데이터를 다운 받고 분석 프로그램을 통해 조합을 할 수 있기 때문에 다양한 조합을 통한 분석이 가능하다. 또한 마커 유전자의 정보가 많을수록 분류도의 신뢰도와 정확성이 커지는 것이 확인되었으므로, 더 많은 마커 유전자의 발견과 해당 유전자 서열의 축적은 NTM 동정 분석에 획기적인 발전을 보일 것으로 기대된다. 구축된 데이터베이스를 이용하여 특정 종이 지니는 유전자 검색과 특정 유전자를 지니는 종 검색을 통해 여러 종이 지니고 있는 유전자를 찾아 낼 수 있을 것이며, 이를 통해 아직 발견되지 않은 잠재적인 NTM의 마커 유전자 발견이 가능할 것으로 기대된다. 잠재적인 마커 유전자를 기존의 NTM종 간에 분류 분석을 통한 마커 유전자로서의 활용도 확인 연구에 이용될 수 있을 것이다.

4.2 연구의 한계점

본 연구의 데이터베이스는 한국에서 연구된 결핵연구원과 제주도대학교에서 연구한 44종의 NTM를 대상으로 구축되었다. 현재까지는 100여종이 넘는 종이 발견되었고 지금도 새로운 종이 발견되고 있기 때문에 이에 대한 나머지 종들에 대한 업데이트가 필요하다. 현재 가장 공신력 있는 NCBI의 GenBank 데이터를 이용하였지만 같은 유전자를 위한 gene 명칭과 product 명칭이 통일되어 있지 않고 불분명하거나 표기가 되어있지 않은 데이터도 있다. 본 연구 결과에서도 NTM 44종중에 사용된 5개의 마커 유전자 중 일부 유전자 서열 정보가 없는 종들이 있었다. 따라서 정기적으로 최신 데이터를 업데이트 하고 각 종별로 추가된 유전자 정보와 함께 마커 유전자 정보도 추가해 나가며, 각각의 근거 자료 검색과 함께 수작업으로 각 유전자에 대한 일관된 명칭을 부여할 예정이다. 본 데이터베이스는 해당 마커 유전자를 지니고 있는 종들과 그 유전자 서열 검색, 한 종이 가지고 있는 다양한 유전자 정보와 서열은 검색할 수 있지만, 복수의 종이 공통으로 지닌 유전자를 알기 위해서는 각각의 종의 유전자 정보를 다운 받고 사용자가 직접 비교하여야 한다. 잠재적인 마커 유전자 발견을 위해서는 복수의 종을 선택하면 그 종들이 공통으로 지닌 유전자가 바로 검색될 수 있는 시스템을 구축할 예정이다. 하지만 이를 위해서는 gene 명칭과 product 명칭을 통일시키는 작업이 선행되어야 하고, 종 명칭 뒷자리에 붙는 strain 부분도 처리되어야 한다는 어려움이 있다. 또한 NTM 종에 대한 시각적 이미지를 포함한 종에 대한 사전 형식의 데이터베이스를 구축함으로써 연구자가 원하는 종의 정보에 대한 접근이 더욱 용이하도록 하여 생화학적 형태 분류학적 정보를 지원하여 유전적 정보를 이용한 계통 분류를 뒷받침하고자 한다. 이처럼 비결핵 마이코박테

리아 동정 연구자에게 필요한 다양한 정보를 지원하여 사용자가 신속하고 정확한 종 동정을 할 수 있도록 웹 데이터베이스를 보완해 나갈 예정이다.

본 연구에서는 다양한 마커 유전자 조합을 이용한 기존 연구들을 토대로 16S rRNA-hsp65-rpoB-sodA-tuf 조합과 16S rRNA-hsp65-rpoB-sodA 조합을 비교하였다. 서열간의 차이를 거리로 나타낸 p-distance를 통해 분별력을 비교하였고, bootstrap을 통해 계통수의 부분별 신뢰도를 평가하였다. 결과적으로 분별력과 견고함에서 더 많은 마커 유전자 조합을 이용한 동정법인 16S rRNA-hsp65-rpoB-sodA-tuf 조합이 우월하였음을 증명하였다. 결과를 살펴보면, 5개 조합 동정법의 p-distance 값이 4개 조합 동정법에 비해 0.001차이로 미세하게 증가하는 것으로 나타났다. 또한 두 조합의 bootstrap 75%이상인 node 비율은 같았지만, 5개 조합의 동정법에서 bootstrap 50%이상인 node가 100%인 반면, 4개 조합의 동정법의 경우는 84.62%이었다. 이처럼 두 동정법이 크지 않은 차이를 보였지만, 분명 신뢰도와 견고함 면에서 5개 동정법이 우월하였음을 수치로 확인할 수 있었다. 이는 앞으로 더 많은 마커 유전자의 발견과 동정법에서의 활용을 통해 더욱 분별력 있고 견고한 동정법 개발이 가능하다는 점을 보여줬다고 생각한다. 본 연구를 통해 균 동정 연구를 통해 얻게 될 다양한 분야에서의 파급 효과를 위해서는 마커 유전자와 동정법 연구가 필요함을 강조하고자 한다.

제 5 장. 결론 및 총론

5.1 결 론

NTM은 의료기술이 발전함에 따라 결핵균과 다르게 분류되었고 인간 환경이 변화함에 따라 결핵균과 대등하게 인간을 위협할 정도로 발병이 급증하고 있다. 결핵균은 인간의 역사와 함께 오랫동안 알려져 왔고 그에 대한 연구가 활발히 진행 되어온 반면, NTM은 인간이 인식한지는 오래되지 않아 그에 대한 연구는 이제 시작 단계이다. NTM은 지리 환경적으로 인간에게 쉽게 노출되어 있을 뿐 아니라 사회적인 문제와도 관련이 있어 우리에게 밀접한 균이라고 볼 수 있다. NTM 연구의 필요성이 커지는 반면 균 특성상 몇 가지 어려움이 있었다. NTM은 결핵균과 분류학적으로 밀접하여 구분이 어렵고 다양한 종이 복합적으로 질병을 일으킨다. 또한 오랜 검진 기간은 신속한 치료를 지연시킨다. NTM은 박테리아 특성상 변이 (mutation)가 빈번하고 종별로 치료법이 다르기 때문에 빠른 진단 및 치료를 위해서는 정확한 유전 정보 파악과 신속하고 효율적인 동정이 필요하다. 이에 본 연구는 NTM에 대한 유전 정보 데이터와 전산 기술을 이용한 신속 정확한 생명정보학적 동정법을 제시하였다. 생명정보학적 동정법은 인간과 더불어 동물, 식물, 균류 등의 유전 서열 정보가 밝혀지면서 생물 연구에서의 활용 가치가 더욱 높아지고 있다. 또한 표현형에 따른 동정법에 비해 신속하고 정확하여 앞으로 균과 관련된 다양한 분야에서 활용될 것으로 기대된다. 이는

병원에서 이루어지는 진단과 치료뿐만 아니라 보다 정확한 NTM 종에 대한 이해로 예방조치 및 관련 의약품의 개발로 이어질 수 있을 것이다. 신속한 NTM 동정은 오랜 검진 기간에 의한 환자의 고통과 진료 비용을 절감해주는 등 상당한 효과를 줄 것이다. 이처럼 NTM 유전 정보와 동정에 대한 연구는 인간의 건강을 증진시키는 보건학적인 측면에서 매우 필요한 연구이다. 그리고 최근 NTM을 포함한 다양한 생명정보 연구에서 시행되고 있는 유전자 조합을 이용한 동정법은 기존에 단일 유전자로 동정할 시 나타나는 한계점을 보완하는 것으로 알려졌다. 따라서 본 연구에서는 44종의 NTM 종과 유전자 서열 정보에 대한 데이터베이스를 구축하여 연구자가 원하는 종과 유전자에 대한 검색이 가능하도록 하였다. 각 종에 대한 유전 정보는 마커 유전자를 포함한 GenBank에 존재하는 모든 유전자에 대한 정보를 담고 있다. 이를 위해 NTM 종과 동정에 이용되고 있는 마커 유전자에 대한 국, 내외 문헌 고찰을 시행하여 NTM 100여종과 10개 정도의 마커 유전자를 파악하였다. 이 중 NCBI의 GenBank에 데이터가 유의하게 존재하고 있는 NTM 44종과 16S rRNA, hsp65, rpoB, sodA, tuf를 이용하여 141095개의 유전 서열 정보를 포함하는 데이터베이스를 구축하였다. GenBank로부터 수집한 데이터는 JAVA language program을 이용하여 NTM 데이터를 필요한 형식으로 추출하였다. DB 형식으로 처리하기 위해 MySQL을 이용하여 저장하였다. 데이터베이스의 웹 서버 구축을 위해 Apache-tomcat을 설치하였고, HTML, Javascript, JSP을 이용하여 연구자 사용을 위한 웹 환경 구축하였다. 이를 통해 사용자가 원하는 NTM 종과 유전자의 정보와 해당 서열 정보를 검색할 수 있고 분석에 활용하도록 다운로드 가능하게 하였다. 이를 본 웹 서버에서 설치 가능한 분석 프로그램을 이

용하여 다양한 마커 유전자 재조합을 통해 다양한 NTM 동정이 가능하다.

본 데이터베이스를 이용하여 마커 유전자 조합을 한 결과, 16S rRNA-hsp65-rpoB-sodA 조합보다는 16S rRNA-hsp65-rpoB-sodA-tuf를 이용한 동정이 분별력과 견고함에서 탁월한 것으로 나타났다. 각각의 단일 마커 유전자의 활용도가 연구를 통해 검증되고, 이러한 마커 유전자들이 많이 발견되어 동정을 위한 조합에 활용될수록 동정법의 신뢰도가 높아지는 것을 알 수 있다. 이를 위해 GenBank와 같은 유전 정보를 관리하는 데이터베이스가 많이 보급되어야 하고 보다 많은 종과 마커 유전자에 대한 정확한 정보 관리가 필요할 것이다. 향후 본 연구에서는 현재 44종과 5개의 마커 유전자에서 추가적으로 보다 많은 NTM 종에 대한 정보와 마커 유전자에 대한 정보를 업데이트할 계획이다. 이로써 마커 유전자 재조합을 통한 NTM 분류 체계 정립에 한 단계 나아갈 수 있는 계기가 될 수 있을 것으로 기대된다.

본 연구는 전산 기술을 이용하여 일반적인 생화학적, 형태분류학적 실험으로 접근할 수 없는 대량의 정보를 다각도로 접근하여 연구자가 다양한 연구를 시행할 수 있는 계기가 되고자 한다. 마커 유전자 정보 파악과 재조합을 통한 종 동정은 NTM에 대한 많은 연구 결과 낳을 것이고 많은 연구 과제를 남길 것으로 기대된다. 이를 통해 최근에 인간의 건강을 위협하고 있는 미지의 균인 NTM에 대한 보다 깊은 이해가 가능하게 할 것이다. NTM에 대한 이해와 연구 성과는 인간을 위협하는 수많은 균에 대한 이해와 함께 인간의 건강 증진이라는 보건학의 최종 목표에 한 걸음 나아갈 수 있는 발판이 될 것이다. 본 연구의 목적은 비결핵 박테리아의 마커 유전자에

기반을 둔 서열 검색, 상동성 비교 및 계통수 분석으로 종 동정에 이용될 수 있는 웹 데이터베이스 지원에 있다. 비결핵 마이코박테리아는 발병률이 꾸준히 증가하며 다양한 질환으로 인간의 건강을 위협하고 있다. 결핵균과 구분이 어렵고 동정하는데 오랜 시간이 소요되는 특징들은 진단과 치료를 더욱 어렵게 한다. 하지만 아직까지 비결핵 마이코박테리아균에 대한 유전자 정보를 담은 오픈 소스의 웹 데이터베이스가 없는 실정이다. 이에 동정이 어려운 비결핵 마이코박테리아균에 대한 유전자와 종 정보를 제공하여 마커 유전자와 종 동정에 대한 연구의 기반이 되고자 한다. 현재까지 꾸준히 비결핵 마이코박테리아균에 대한 동정 연구가 진행되어오고 있지만 분별력 있고 확고한 분류 체계를 나타내는 단일 유전자는 아직 밝혀지지 않고 있다. 하지만 비결핵 마이코박테리아균은 상호간에 유전적 상동성을 지니고 있고 단일 마커 유전자만으로는 확고한 분류 체계를 정의할 수 없다고 생각된다. 이에 현재까지 마커 유전자인 16S rRNA, hsp65, rpoB, sodA, tuf 이외에 더 많은 종을 아우르고 종간의 분별력을 나타내는 새로운 마커 유전자의 연구가 필요하고 그러한 유의미한 마커 유전자들의 조합을 통해 분별력 있고 확고한 분류 체계 확립이 가능할 것이다. 본 데이터베이스의 향후 보완을 통해 전 세계적으로 알려진 100여 종 이상의 비결핵 마이코박테리아 종에 대한 유전 정보를 업데이트하고 아직 계통 분류학적인 분석이 이루어지지 않은 새로운 종에 대한 연구에 활용 될 수 있다. 지금까지 발견된 모든 종에 대한 계통 분류학적 분석을 통해 인간에 특정 질병을 일으키는 균과 발병 및 균의 생명 활동에 관여하는 유전자들을 파악하는 연구에 활용될 수 있다. 이러한 웹 데이터베이스가 갖춰지면 새로운 비결핵 마이코박테리아균 감염 질환이 발견되었을

경우 신속한 종 동정과 치료를 보장할 수 있다. 이는 진단과 치료법 연구를 거쳐 특정 균의 특정 유전자를 타깃으로 하는 의약품 및 부작용 없이 질병 관련 균만 제거하는 치료법 개발에 활용 될 것이다. 이러한 연구는 균의 유전자와 분류 체계의 정확한 파악에서 비롯되어야 한다. 이는 질병 예방과 신속한 질병 확산 방지, 치료로 이어지면서 보건학적 연구 활용에 큰 보탬이 되리라 기대된다. 비결핵 마이코박테리아균이 인간 거주지에 널리 퍼져있듯이 대부분의 균은 인간 생활과 밀접한 관련이 있다. 인간이 인지하지 않더라도 꾸준히 다양한 균의 영향을 받아 왔다. 균에 대한 생물학적 이해는 인간 생활을 건강하고 윤택하게 할 것이고 이를 위해 동정 연구가 가장 필수적이다. 이러한 연구는 지금까지 인간의 질병과 관련된 균에 대한 동정 연구와 앞으로 새롭게 발견되는 다양한 균에 대한 동정 연구에 초석이 되어 질병 진단 및 치료뿐만 아니라 식품, 농수산물, 항공 우주 관련 산업에서 균 동정 및 배양에 대한 연구에도 활용될 것이다.

구축된 데이터베이스를 다시 한 번 살펴보면 “NTM gene search” 페이지에서 검색을 통해 accession을 확인하여 “Marker-sequence searching” 페이지에서 서열 정보를 검색하여 다운 받거나 직접 organism, gene, 또는 product 명을 검색하여 다운로드 할 수 있다. “Blast & Analysis Programs” 페이지에서는 ClustalX, SequenceMatrix, MEGA 프로그램을 링크하여 연구자가 다운로드한 데이터를 이용하여 직접 서열 정렬, 조합, 분석할 수 있도록 하였다. 또한 NCBI의 BLAST 웹 링크를 통하여 연구자가 상동성 검색이 필요할 경우 사용할 수 있도록 하였다. GenBank의 데이터를 기반으로 한 데이터베이스의 유전 정보는 기존에 알려진 마커 유전자 이 외에는 각 종마

다 다양한 유전자를 보유하고 있었다. 이는 유의미한 분류 체계를 위한 새로운 마커 유전자 발견을 위해 아직까지 연구해야 할 영역이 많이 남아있음을 알 수 있었다. NTM에 대한 유전 정보 연구가 새로운 마커 유전자 연구를 위해서는 더 많은 연구와 분석이 이루어져야 함을 알 수 있었다. 본 연구에서 개별 마커 유전자와 복합 유전자 조합의 계통수 분석을 실시하였다. 개별 마커 유전자에 의한 계통수 분석은 각 유전자에 따라 조금씩 결과가 다르게 나타나서 특정 유전자만을 이용한 분석에 한계가 있었다. 같은 맥락으로 각 마커 유전자의 WWWBlast를 이용한 상동성 분석 결과에서도 확인할 수 있었다. 그 다음으로 4개 유전자 조합과 5개 유전자 조합에 의한 분류 체계법을 확인하였다. 먼저 복합 유전자 조합에 의한 계통수 분석이 개별 유전자 계통수 분석보다 신뢰도 있는 분석법임을 확인할 수 있었다. 그리고 4개 와 5개 유전자 조합 분류 체계를 비교한 결과 분별력과 신뢰도에서 5개 조합을 이용한 분류 체계가 우월함을 확인 하였다. 본 연구는 더 많은 마커 유전자를 활용한 동정법이 향상된 분별력과 견고함을 나타냄을 보여주고 앞으로 더 많은 마커 유전자 발견과 동정법 연구가 필요함을 증명하였다. 단순히 조합 유전자 개수를 늘리는 것보다는 유의미한 마커 유전자의 정밀한 선별 작업에 선행된 조합을 형성해야 함을 것을 간과해서는 안 될 것이다. 본 데이터베이스는 모든 NTM 종에 대한 유전 정보를 담은 데이터베이스 구축을 목표로 향후 보완 절차가 이루어질 것이다. 앞으로 새롭게 발견되는 NTM 종과 유전 정보를 분석할 수 있는 밑거름이 될 것이고 새로운 마커 유전자를 발견하기 위한 연구에 보탬이 될 것으로 기대된다. 이는 NTM종뿐만 아니라 다른 균류의 공유전 정보나 분류 체계에 대한 연구에 기반이 될 것으로 기대된다.

결론적으로 본 연구는 생명정보학적 데이터 분석을 통해 기존의 실험적인 NTM 분석법에서 발견할 수 없었던 유전 연구의 새로운 방법론을 제시할 것이다. 또한 인간의 생활에 영향을 미치는 균에 대한 다양한 활용으로 농산물 개량, 식품개발과 같은 다양한 산업 분야에 영향을 미칠 것으로 기대된다. 특히 인간의 건강에 영향을 미치는 균에 대한 유전적 이해는 질병 예방, 진단, 치료 모든 분야에 효율성을 높임으로써 보건학적 의미 또한 상당할 것으로 기대된다. 특히 생명정보학적 기술의 보건학적 이용은 앞으로 보건학에서 생명정보학 기술과 방법론이 널리 사용되어야 할 당위성을 보여주고 있다.

5.2 총 론

비결핵 마이코박테리아균(NTM)은 마이코박테리아 속(genus)에서 결핵균과 나병균을 제외한 균을 말한다. 비결핵 항산균 또는 비정형 항산균이라고도 불리고 100여종이 넘는 균종이 있다. 비결핵 마이코박테리아균은 1980년 전까지 결핵균에 비해 위험성이 덜 인식되어 오다가 AIDS의 확산으로 인해 크게 확산되었다. 나라별 구체적인 현황은 조금씩 다르지만 전반적으로 결핵균 질환 발병률이 감소하는 반면에 비결핵 항산균 질환 발병률이 증가하여 최근에는 비슷한 발병률을 나타내고 있다. 결핵균과 비결핵 마이코박테리아균을 분류하는 생화학적 방법으로 항산균도말검사, 배양검사, 핵산증폭검사가 있다. 결핵균이 특정 환경에서만 발견되는 반면, 비결핵 박테리아균은 강이나 논가와 같이 인간의 생활 주거지와 밀접한 환경에서 발견된다. 관련 질환으로는 호흡기, 림프절염, 피부염, 크롬병, 관절염 등이 있으며 이 중 90% 이상이 호흡기 질환에 해당한다. 대표적인 호흡기 질환을 일으키는 비결핵 마이코박테리아균은 *M. kansasii*, *M. avium*, *M. intracellulare* 등이 있다. 그 외에 림프절염 원인균으로 MAC, *M. scrofulaceum*, 피부염 원인균으로 *M. szulgai*, *M. ulcerans*, *M. marinum* 등이 있다. 우리나라에서 가장 흔한 균은 MAC로 비결핵 마이코박테리아균 발견의 60~80%를 차지한다. 비결핵 마이코 박테리아 질환의 진단 기간은 수개월에서 수년이 걸리기도 하고 감염균이 결핵균과 구분이 어려워 오진이 발생하는 경우도 많기 때문에 치료와 진단에 어려움이 있다.

NTM 분류는 마이코박테리아 분류법인 Runyon classification에 의해 이루어진다. 증식 속도에 따라 10일 전 후로 신속발육균과

완속발육균으로 구분하고 완속발육균은 색소생성에 따라 비발색균, 광발색균, 암발색균으로 분류한다. NTM 질병의 확산은 1990년대 유전자 서열 분석의 발전과 더불어 NTM 종의 동정법에 마커 유전자 서열 정보 분석을 불러일으키게 되었다. NTM의 대표적 마커 유전자는 16S rRNA, hsp65, rpoB, sodA, tuf 등이 있다. 각 유전자를 이용한 동정법은 조금씩 다른 결과를 나타내고 변별력과 견고함에 있어 부족하다. 최근에는 다양한 마커 유전자를 조합하여 한 가지 유전자만을 이용한 분석의 한계점을 극복하고 있다. 국내에서는 결핵 연구원, 제주대학교, 서울대학교 등에서 NTM에 대해 연구하고 있고 해외 연구 기관에서 결핵균에 대한 검색 DB를 구축하기도 하였지만 아직까지는 NTM에 대한 오픈 소스의 데이터베이스가 없는 실정이다. 동정이 어려운 NTM균을 더욱 신속하고 빠르게 동정하기 위해 공통된 분석 데이터베이스의 구축이 절실히 필요하다. 또한 연구자가 마커 유전자 조합을 이용하여 동정이 가능하도록 한다면 더욱 변별력 있고 견고한 분류 체계 분석이 가능할 것이다. 본 연구의 목적은 NTM 종에 대한 유전 정보, 마커 유전자의 서열 정보를 담고 있는 데이터베이스를 구축하여 사용자가 원하는 정보를 검색하고 해당 정보를 다운 받아 관련 프로그램들을 이용할 수 있도록 하고자 한다. 사용자는 ClustalX와 SequenceMatrix 프로그램을 이용하여 복수의 NTM 마커 유전자를 정렬 및 조합할 수 있다. 정렬 조합된 서열은 MEGA를 이용하여 계통수 분석을 포함한 다양한 분석 연구가 가능하도록 하였다. 본 데이터베이스를 통해 각 NTM 연구기관과 연구자들이 같이 연구하고 정보를 공유할 수 있는 터전이 되고 다양한 균 동정법에 생명정보학적 분석 기법이 이용되도록 하는 것이 본 연구의 목적이다. 최종적으로 본 데이터베이스에 의한 NTM 종 분

석법은 인간의 건강 증진을 위한 보건학적 수단이 될 것으로 기대된다.

본 연구에서 사용된 데이터는 NCBI의 GenBank ftp 사이트에서 다운로드 하여 수집하였다. 142개의 gbbct 파일을 JAVA 프로그래밍 언어를 이용하여 데이터를 추출하였다. gbbct 파일은 organism, definition, version, reference, gene, region 등 종과 유전자에 대한 자세한 정보를 담고 있다. 이를 JAVA 언어를 이용하여 데이터베이스 구축에 필요한 종과 그에 대한 데이터를 추출하였다. 웹 서버의 데이터베이스 테이블 입력을 위해 accession, organism, definition, CDS, gene, product의 각 항목을 추출하여 탭(tab)으로 구분하였다. accession은 서열 고유 번호, organism은 종명, definition은 종과 서열에 대한 구체적인 설명, CDS는 coding DNA sequence 부분의 좌표, gene과 product는 각각 유전자 이름과 유전자 산물을 나타낸다. 이를 검색 가능한 MySQL 데이터베이스에 파일 형식으로 저장하여 사용자가 구체적인 수열 정보를 검색할 수 있도록 하였다. 분석을 위한 종과 유전자의 서열 정보는 추출하여 분석이 용이하도록 fasta 형식으로 저장하였다. NTMDB 웹 데이터베이스는 메인 페이지를 포함해서 총 5가지 페이지로 구성되어있다. 메인 페이지는 NTM과 본 웹 데이터베이스에 대한 간단한 설명으로 구성되어있다. “NTM Gene Search” 페이지에서는 원하는 종의 정보를 검색하여 확인할 수 있다. “Marker-sequence Search”페이지는 유전자 서열 검색을 통해 서열 정보를 다운로드 받을 수 있다. “Sequence Fasta Files” 페이지는 사용자가 검색하지 않고 간편하게 44종과 5개 주요 마커 유전자에 대한 서열 정보를 다운로드 할 수 있도록 fasta file을 제공하고 있다. “Blast & Analysis Programs” 페이지는 동정 분석에 필요한 ClustalX, SequenceMatrix,

MEGA 프로그램을 제공하고 있다.

본 연구에서는 NTMDB를 이용하여 추출한 16S rRNA, hsp65, rpoB, sodA, tuf 각각의 마커 유전자를 이용한 계통 분류도를 제작하고 기존 연구와의 비교를 통해 각 마커 유전자의 활용도를 확인하였다. 그 후에 기존에 연구되었던 16S rRNA, hsp65, rpoB, sodA를 조합한 계통 분류도와 16S rRNA, hsp65, rpoB, sodA, tuf까지 포함한 계통 분류도를 작성하고 비교 분석하였다. 이를 통해 NTM 연구에 다양한 마커 유전자 조합이 이용될 필요성에 대해 제시하고자 하였다. 분석 방법은 Neighbor-joining 방식의 분류 계통도에 kimura-2-parameter 모델에, bootstrap replication은 500으로 설정하였다. 16S rRNA 서열을 이용한 계통 분류도는 본 44종의 NTM 데이터베이스 데이터 중 기존 연구(Devulder et al., 2005)와 공통인 33종을 추려 계통 분류하였다. 기존 연구와 비교한 결과, 일부 종들을 제외하고 전반적으로 NTM이 발육 속도에 따라 위쪽은 완속발육군, 아래쪽은 신속발육군으로 분류가 이루어진 것을 볼 수 있다. *M. gilvum*, *M. parafortuitum*; *M. ulcerans*, *M. marinum*; *M. avium*, *M. paratuberculosis*; *M. chelonae*, *M. abscessus*의 유전적 연관성을 나타낸 것도 확인할 수 있었다. Bootstrap 치수는 0개의 node 중 bootstrap 치수가 75 이상인 node는 8개, 50 이상인 node는 18개였다. 16S rRNA 유전자에 의한 계통 분류도에서 *M. paratuberculosis*, *M. avium*와 tuf 유전자에 의한 계통 분류도에서 *M. abscessus*, *M. bolletii*와 같이 분류적 연관성은 있으나 구분하지 못하는 경우도 있었다. 이는 다섯 유전자가 지금까지 여러 연구에서 마커 유전자로 입증되어 왔지만, 어느 하나만의 유전자로는 정확도에 일관성 있는 계통 분류가 어렵다는 것을 알 수 있었다.

16S rRNA, hsp65, rpoB, sodA를 조합한 계통 분류도와 16S

rRNA, hsp65, rpoB, sodA, tuf를 조합한 계통 분류도를 비교하여 조합에 따라 분류 계통도의 분별력과 견고함을 확인하였다. 16S rRNA, hsp65, rpoB, sodA, tuf의 종들 중 공통으로 보유하고 있는 16종을 대상으로 유전자를 ClustalX를 이용하여 서열 정렬 후 SequenceMatrix를 이용하여 유전자를 조합하였고 MEGA를 통해 계통 분류도를 작성하였다. 두 분류도를 비교한 결과 발육 속도, 종들 간의 연관성이 거의 일치하였고 bootstrap 치수에서는 약간의 차이를 보였다. 16S rRNA-hsp65-rpoB-sodA 조합의 경우 50 이상의 bootstrap 치수를 나타내는 node가 13개 중 11개를 나타낸 것에 비해, 16S rRNA-hsp65-rpoB-sodA-tuf 조합에서는 13개의 node 중 13개 모두 50 이상의 bootstrap 치수를 보였다. 반면에 75 이상의 bootstrap 치수는 두 조합 모두 13개 node 중 10개로 나타났다. 이는 50개 이상의 node를 더 많이 보유한 16S rRNA-hsp65-rpoB-sodA-tuf의 조합이 조금 더 견고한 계통 분류도를 나타내고 있음을 나타낸다. 평균 p-distance와 평균 kimura-2-parameter model distance를 보면 각 유전자의 수치는 제 각각이지만 16S rRNA-hsp65-rpoB-sodA 조합과 16S rRNA-hsp65-rpoB-sodA-tuf 조합은 각각 0.001씩 차이를 나타내어 16S rRNA-hsp65-rpoB-sodA-tuf 조합이 계통 분류적인 분별력이 상승된 것을 알 수 있었다. 또한 두 조합의 bootstrap 수치에서는 75% 이상의 비율은 같지만 50% 이상의 비율은 16S rRNA-hsp65-rpoB-sodA-tuf 조합이 100%이기 때문에 계통 분류도의 견고함이 상승된 것을 볼 수 있었다. 결과적으로 두 조합의 P-distance와 K2P distance, 그리고 bootstrap을 비교하면 16S rRNA-hsp65-rpoB-sodA-tuf 조합이 16S rRNA-hsp65-rpoB-sodA 조합에 비해 변별력과 견고함에 있어서 상대적으로 더 안정된 계통 분류도 모델임을 알 수 있다. 본 연구에서는 NTM 44종의 16S rRNA,

hsp65, rpoB, sodA, tuf 염기 서열에 대한 데이터베이스를 구축하여 *M. peregrinum*을 기준으로 blastn과 tblastx 분석으로 상동성 검색을 실시하였다. 대부분의 분석 결과에서 *M. peregrinum*과 가장 유사한 종에 대해서 일치하였지만 그 이하 진화적으로 가까운 종들에 대해서는 조금씩 다른 양상을 보였다. 이와 같이 종 분류에 대한 연구는 어느 한 가지 분석법에 의존 할 수 없고 더 많은 분석법이 개발될 필요가 있고 이를 종합하여 해석해야 함을 알 수 있었다.

본 연구의 기대성과는 학술적으로 기존의 실험적이고 형태학적인 NTM 분석법의 한계점을 극복하고 생물학적 이론과 첨단 컴퓨터 기술이 결합한 생명정보학적 분석을 통해 NTM 연구에 접목시킴으로써 새로운 연구 분야의 길을 제시할 것이다. 산업적인 측면에서는 의학, 보건학, 약학과 같이 건강과 직결된 산업 분야뿐 아니라 의류업, 식품업, 군수업, 농수산업 같이 일반 생활과 밀접한 연관이 있는 전반적인 산업 분야에 걸쳐 NTM과 같은 박테리아균의 유전적 정보는 크게 영향을 미칠 것으로 기대된다. 기술적 측면에서는 생명정보학적 분석법은 기존의 분석법과 다르게 다각도로 무궁무진한 접근이 가능하고 인간으로서는 불가능한 대용량 정보 처리를 컴퓨터를 이용하여 실행되기 때문에 컴퓨터 기술의 발전과 함께 앞으로 도 무한한 발전 가능성을 제시할 것이다. 마지막으로 본 연구는 인간의 건강을 최선의 목표로 하는 보건학적인 측면에서 활용도가 매우 클 것으로 기대된다. NTM의 빠른 종 동정은 진단 기간을 단축하여 신속한 환자의 치료로 이어져 환자의 건강을 지키는 보건학적인 역할을 수행할 수 있다. 지금까지 비결핵 마이코박테리아균에 대한 오픈 소스의 웹 데이터베이스가 없는 실정에서 본 데이터베이스는 향후 보완을 통해 지금까지 밝혀진 100여종의 NTM에 대한 정보를

업데이트하여 더 많은 종을 아우르고 종간의 분별력을 나타내는 새로운 마커 유전자의 연구와 확고한 분류 체계 연구에 활용될 것이다. 아직 계통 분류가 이루어지지 않은 새로운 종에 대한 연구에 활용될 수 있을 것이다. 또한 새로운 비결핵 마이코박테리아균 감염 질환이 발견되었을 경우 신속한 종 동정과 치료를 보장할 수 있다. 이는 진단과 치료법 연구를 거쳐 특정 균의 특정 유전자를 타겟으로 하는 의약품 및 부작용 없이 질병 관련 균만 제거하는 치료법 개발에 활용될 것이다. 이러한 연구는 균의 유전자와 분류 체계의 정확한 파악에서 비롯되어야 한다. 이는 질병 예방뿐 아니라 신속한 질병 확산 방지와 치료로 이어지면서 보건학적 연구 활용에 큰 보탬이 되리라 기대된다. 인간의 질병과 관련된 균에 대한 동정 연구, 질병 진단 및 치료, 그리고 식품, 농수산물, 항공 우주 관련 산업에서 균 동정 및 배양에 대한 연구에도 활용될 것이다.

참 고 문 헌

- Adekambi T, Berger P, Raoult B, Drancourt M. rpoB gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium aubagnense* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*. 56(1):133–143 (2006).
- Adekambi T, Drancourt M. Dissection of Phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, hsp65, sodA, recA and rpoB gene sequencing. *International Journal of Systematic and Evolutionary Microbiology*. 54:2095-2105 (2004).
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol.* 215(3):403-410 (1990).
- Alcaide F, Richter I, Bernasconi C, Springer B, Hagenau C, Schulze-Robbecke R, Tortoli E, Martin R, Bottger EC, Telenti A. Heterogeneity and Clonality among isolates of *Mycobacterium kansasii*: implications for epidemiological and pathogenicity studies. *J. Clin. Microbiol.* 35(8):1959-1964 (1997).
- Andrejak C, Thomsen VO, Johansen IS, Riis A, Benfield TL, Duhaut P, Sorensen HT, Lescure F, Thomsen RW. Nontuberculous Pulmonary Mycobacteriosis in Denmark- Incidence and Prognostic Factors. *American Journal of Respiratory and Critical Care Medicine* 181(5):514-521 (2010).
- Blackwood KS, He C, Gunton J, Turenne CY, Wolfe J, Kabani AM. Evaluation of recA Sequences for Identification of *Mycobacterium* Species. *J. Clin. Microbiol.* 38(8):2846-2852 (2000).
- Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies. *Appl. Environ. Microbiol.* 73(1):278-288 (2007).
- Cheon J, Lee M, Ki J. Molecular Divergences of 16S rRNA and rpoB Gene in Marine Isolates of the Order Oscillatoriales (Cyanobacteria). *Korean Journal of Microbiology*. 48(4):319-324 (2012).
- David HL. Carotenoid Pigments of *Mycobacterium kansasii*. *Applied Microbiology*. 28(4):696-699 (1974).

- Devulder G, de Mpntclos P, Flandrois JP. A multigene approach to phylogenetic analysis using the Genus *Mycobacterium* as a model. *International Journal of Systematic and Evolutionary Microbiology*. 55(1):293–302 (2005).
- Falkinham III JO. The Changing pattern of nontuberculous mycobacterial disease. *Can J Infect Dis*. 14(5):281-286 (2003).
- Falkinham III JO. Epidemiology of Infection by Nontuberculous Mycobacteria. *Clinical Microbiology Review*. 9(2):177-215 (1996).
- Filer D, Furano AV. Duplication of the *tuf* gene, which encodes peptide chain elongation factor Tu, is widespead in Gram-negative bacteria. *J. Bacteriol*. 148(3):1006-1011 (1981).
- Filer D, Furano AV. Portions of the Gene Encoding Elongation Factor Tu Are Highly Conserved in Prokaryotes. *The Journal of Biological Chemistry*. 255(2): 728-734 (1980).
- Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Fordin F, Holland SM, Horsburgh R, Ultt G, Lademarco MF, Iseman M, Olivier K, Ruoss S, Reyn CF, Wallace RJ Jr, Winthrop K. An Official ATS/IDSA Statement: Diagnosis, Treatment, and Prevention of Nontuberculous Mycobacterial Diseases. *Am J Respir crit Care Med*. 175(4):367-416 (2007).
- Hirano K, Abe C, Takahashi M. Mutations in the *rpoB* Gene of Rifampin-Resistant *Mycobacterium tuberculosis* Strains Isolated Mostly in Asian Countries and Their Rapid Detection by Line Probe Assay. *J. Clin. Microbiol*. 37(8):2663-2666 (1999).
- Jo JM, Shin MS, Kim JH, Kim MJ, Park HJ, Na HK, Jo KU, Kim J, Shim TS. Two Cases of Nontuberculous Mycobacterial Lung Disease in Heart Transplant Recipients. *Tuberculosis and Respiratory Diseases*. 69(3):201-206 (2010).
- Kasai H, Ezaki T, Harayama S. Differentiation of Phylogenetically Related Slowly Growing Mycobacteria by Their *gyrB* Sequences. *J. Clin. Microbiol*. 38(1):301-308 (2000).
- Katoch V.M. Infections due to non-tuberculous Mycobacteria (NTM). *Indian J Med Res* 120:290-304 (2004).
- Kim B, Lee S, Lee K, Park C, Choi M, Kim I, Choi S, Hwang E, Cha C, Kim S, Bai G, Kook Y. Identification of *Mycobacterium avium* complex (MAC) Clinical Strains to a Species Level by Sequencing and PCR-SSCP Analysis of *rpoB* DNA. *J. Korean Soc. Microbiol.*, 34(5):491-500 (1999).

- Kim B, Lee S, Lyu M, Kim S, Bai G, Kim S, Chae G, Kim E, Cha C, Kook Y. Identification of Mycobacterial Species by Comparative Sequence Analysis of the RNA Polymerase Gene (rpoB). *J. Clin. Microbiol.* 37(6):1714-1720 (1999).
- Kim H, Kim S, Shim T, Kim M, Bai G, Park Y, Lee S, Chae G, Cha C, Kook Y, Kim B. Differentiation of Mycobacterium species by analysis of the heat-shock protein 65 gene (hsp65). *International Journal of Systematic and Evolutionary Microbiology.* 55(4):1649–1656 (2005).
- Kimura M. A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequences. *J. Mol. Evol.* 16: 111-120 (1980).
- Kim T, Kim H, Kwon S, Park H. Analysis of Bacterial Community Composition in Wastewater Treatment Bioreactors Using 16S rRNA Gene-based Pyrosequencing. *The Korean Journal of Microbiology.* 46(4):352-258 (2010).
- Koh W, Kwon J. Diagnosis and treatment of nontuberculous mycobacterial lung disease. *The Korean Journal of Medicine.* 74(2):120-131 (2008).
- Korea Center for Disease Control & Prevention. KOREAN GUIDELINES FOR TUBERCULOSIS, 1st edition. (2011).
- Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. 10(2):189-191 (1994).
- Lai C, Tan C, Chou C, Hsu H, Liao C, Huang Y, Yang P, Luh K, Hsueh P. Increasing Incidence of Nontuberculous Mycobacteria, Taiwan, 2000-2008. *Emerging Infection Diseases.* 16(2):294-296 (2010).
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Cluster X version 2.0. *Bioinformatics.* 23(21):2947-2948 (2007).
- Lee SD, Lee HY, Kim HC, Kim SY. Mycobacterium tuberculosis and Nontuberculous Mycobacteria by PCR Assay. *Korean J Clin Microbiol.* 10(2): 135-142 (2007).
- McCord JM, Fridovich I. Superoxide Dismutase-An Enzymic Function for Erythrocyte (Hemocytin). *The Journal of Biological Chemistry.* 244(22): 6049-6055 (1969).
- McNabb A, Eisler D, Adie K, Amos M, Rodrigues M, Stephens G, Black WA, Isaac-Renton J. Assessment of Partial Sequencing of the 65-Kilodalton Heat Shock Protein

Gene (hsp65) for Routine Identification of Mycobacterium Species Isolated from Clinical Sources. *J. Clin. Microbiol.* 42(7):3000-3011 (2004).

Melov S, Doctrow SR, Schneider JA, Haberson J, Patel M, Coskun PE, Huffman K, Wallace DC, Malfroy B. Lifespan Extension and Rescue of Spongiform Encephalopathy in Superoxide Dismutase 2 Nullizygous Mice Treated with Superoxide Dismutase-Catalase Mimetics. *The Journal of Neuroscience.* 21(21):8348-8353 (2001).

Mignard S, Flandrois J. Identification of Mycobacterium using the EF-Tu encoding (tuf) gene and the tmRNA encoding (ssrA) gene. *Journal of Medical Microbiology.* 56(8):1033-1041 (2007).

Noh E, Kim Y, Kim D, Kim K. Bacterial Diversity in the Guts of Sea Cucumbers (*Apostichopus Japonicus*) and Shrimps (*Litopenaeus vannamei*) Investigated with Tag-Encoded 454 Pyrosequencing of 16S rRNA Genes. *Korean Journal of Microbiology.* 49(3):237-244 (2013).

Primm TP, Lucero CA, Falkinham III JO. Health Impacts of Environmental Mycobacteria. *Clin. Microbiol.* 17(1):98-106 (2004).

Rogall T, Wolters J, Flohr T, Bottger EC. Towards a Phylogeny and Definition of Species at the Molecular Level within the Genus Mycobacterium. *International Journal of Systematic Bacteriology.* 40(4):323-330 (1990).

Saitou N, Nei M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 4(4):406-425 (1987)

Sela S, Yogeve D, Razin S, Bercovier H. Duplication of the tuf gene: a new insight into the phylogeny of eubacteria. *J. Bacteriol.* 171(1):581-584 (1989).

Shin JH, Cho EJ, Lee JY, Yu JY, Kang YH. Novel Diagnostic Algorithm Using tuf Gene Amplification and Restriction Fragment Length Polymorphism is Promising Tool for Identification of Nontuberculous Mycobacteria. *J. Microbiol. Biotechnol.* 19(3):323-330 (2008).

Singh S, Gopinath K, Shahdad S, Kaur M, Singh B, Sharma P. Nontuberculous Mycobacterial Infections in Indian AIDS Patients Detected by a Novel Set of ESAT-6 Polymerase Chain Reaction Primers. *Jpn. J. Infect. Dis.* 60(1):14-18 (2007).

Springer B, Bottger EC, Kirschner P, Wallace RJ Jr. Phylogeny of the Mycobacterium chelonae-like Organism based on Partial Sequencing of the 16S rRNA Gene and Proposal of Mycobacterium mucogenicum sp. nov. *International Journal of Systematic Bacteriology.* 45(2):262-267 (1995).

Takewaki S, Okuzumi K, Ishiko H, Nakahara K, Ohkubo A, Nagai R. Genus-specific polymerase chain reaction for the mycobacterial dnaJ gene and species-specific oligonucleotide probes. J. Clin. Microbiol. 31(2):446-450 (1993).

Torriani S, Felis GE, Dellaglio F. Differentiation of *Lactobacillus plantarum*, *L. pentosus*, and *L. paraplantarum* by recA Gene Sequence Analysis and Multiplex PCR Assay with recA Gene-Derived Primers. Appl. Environ. Microbiol. 67(8):3450-3454 (2001).

Tortoli E. Impact of Genotypic Studies on Mycobacterial Taxonomy: the New Mycobacteria of the 1990s. Clin. Microbiol. 16(2):319-354 (2003).

Tsukamura M. Classification of Scotochromogenic Mycobacteria. Japan. J. Microbiol. 12(1):63-75 (1968).

Vaidya G, Lohman DJ, Rudolf M. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 27(2):171-180 (2011).

Yim J, Park Y, Lew WJ, Bai G, Han SK, Shim Y. Mycobacterium Kansaii Pulmonary Diseases in Korea. J Korean Med Sci. 20(6):957-960 (2005).

유희경, 심명섭, 양정성, 박영길, 김창기, 김희진. 우리나라 임상환자에서 분리된 비결핵마이코박테리아. 제112차 대한결핵 및 호흡기학회 추계학술대회. TP-60 (2011).

Abstract

Non-tuberculous Mycobacteria Classification System Based On Multiple Marker Genes

Cho, Kwang Hoon

**Laboratory of Computational Biology and Bioinformatics
Graduate school of Public Health
Seoul National University**

There are several bacteria which are helpful or harmful for human beings in the environment. Non-tuberculous mycobacteria (NTM) are the species belong to Mycobacterium genus other than Mycobacterium leprae and Mycobacterium tuberculosis. Non-tuberculous mycobacteria had been spread out with AIDS breakout in 1980s. Non-tuberculous mycobacterium exists in the environments such as water and land while mycobacterium tuberculosis does not. Non-tuberculous mycobacterium causes lymphadenitis, skin, pulmonary diseases, and 90% of NTM diseases are pulmonary. Lymphadenitis is caused by Mycobacterium avium complex (MAC). M. scrofulaceum, and disseminated disease is caused by MAC, M. kansasii, M. scrofulaceum, and M. xenopi. Joint disease is caused by M. non-chromogenicum, and M. kansasii. Crohn's disease is caused by M. paratuberculosis. While mycobacterium tuberculosis diseases have been decreased by the research of diagnosis and treatment, non-tuberculous mycobacterium diseases have been increased in all over the countries and

regions. The rates of mycobacterium tuberculosis diseases and non-tuberculous mycobacterium diseases have become similar recently. Among over 100 NTM species discovered so far the most common non-tuberculous mycobacterium species in Korea is mycobacterium avium complex, which takes 60~80%. *M. abscessus* is secondly common in Korea but rare in other countries. *M. kansasii*, otherwise, is secondly common cause of pulmonary disease in the U.S and Japan, but it is uncommon in Korea. Non-tuberculous mycobacterium is identified by AFB smear, sputum examination, and PCR analysis. However, biochemical examination takes several months or years to identify, and some causes the confusion between non-tuberculous mycobacterium and mycobacterium tuberculosis. Therefore, there should be the advanced method for species classification with bioinformatics technique. Several recent research show species classification methods using multiple marker gene, and these provides more sensibility and robustness than using one marker gene. This database has been constructed for supporting the genetic data for non-tuberculous mycobacterium species and their marker genes. By using NTM genetic data from this database, researchers can combine marker gene sequence in several ways, and could result in more sensible NTM classification. This database provides the function of searching and downloading 44 species NTM genetic data and 16S rRNA, hsp65, rpoB, sodA, and tuf marker gene data. 16S rRNA and hsp65 are the most suitable marker genes for NTM analysis. 16S rRNA is for the all of prokaryotes, and hsp65 is the antigen protein related in the immune system for NTM species. RpoB has the genetic information on β -subunit of bacterial RNA polymerase. Superoxide dismutase (SOD) is the gene for the enzyme activating detoxication. Tuf gene is for the elongation factor EF-Tu expression. As tuf gene repeated in gram negative bacteria, it is used in the bacterial marker gene analysis. The list of 44 NTM species is selected from the previous research,

and the gene data of NTM species is collected from GenBank gbbct files through ftp protocol. Parsing the data and converting it to the fasta format are executed using JAVA language. The filtered data is loaded on web server using MySQL, HTML, javascript, JSP, and apache-tomcat. The web server provides ClustalX, MEGA, and SequenceMatrix, WWWBlast programs to execute the local-blast, alignment, and combination of NTM multiple marker genes. At the result, this study proves that 5 marker gene classification method using 16S rRNA-hsp65-rpoB-sodA-tuf is more sensible and robust than 4 marker gene classification method using 16S rRNA-hsp65-rpoB-sodA from the p-distance and bootstrap analysis. This study clarify that the development of genetic data searching system is very important. This study suggests the novel way for diagnosis and treatment for NTM and so as other bacteria. This study would be meaningful for the purpose of public health field.

.....
Keywords: Non-tuberculous Mycobacterium, NTM, Bioinformatics,
Database, Phylogenetics, Marker Gene, Public Health

Student ID: 2011-23864